

A fresh start in space

Whatever their previous scepticism, scientists should embrace the opportunity to see whether the International Space Station can address important research questions. NASA's decisions on its microgravity research are a step in the right direction.

Sixteen years after it was first proposed — and the wait has seemed every bit that long — the International Space Station is finally a reality. The first crew moved in last week, and one of its first acts, appropriately, was to give the new outpost a name.

US astronaut William Shepherd, who, along with two Russian cosmonauts, will live on the station for four months, had long lobbied for a less unwieldy name than 'International Space Station'. So during a televised chat with Daniel Goldin, chief of the US space agency NASA, Shepherd asked if the astronauts could call their new home 'Alpha'. Goldin, probably fearing a political wrangle among the station's international partners, had previously resisted any such move, but this time he gave in. So 'Alpha' it is — at least while Shepherd's crew is living there.

Purists will point out that this is not humanity's first space station. The Russians orbited the first Salyut station almost 30 years ago, followed by the US Skylab and the Russian Mir. But none of those projects approached the new station in terms of size, complexity or ambition. So even if Alpha is not an original idea, it can justifiably be called a fresh start.

NASA, too, seems poised to begin a new era of scientific research in orbit. The recent reorganization of the agency's microgravity and life-science programme (see page 123) bodes well for a more mature outlook on what can and cannot be accomplished in space. Gone are the grandiose claims, which raised eyebrows in the past, about space-based experiments leading to cures for cancer and AIDS. The focus

has shifted away from the dubious accomplishments of past crystal-growth experiments on the space shuttle — once touted as a commercial bonanza for pharmaceutical companies but discredited by those who showed that it was cheaper to grow such crystals on Earth — and towards solid, peer-reviewed studies of gravitational biology.

Scientists involved in this highly specialized area of research have much to look forward to. Their prospects include far better laboratory facilities than those on board the shuttle or Mir, and the chance to repeat experiments and continue them for long periods. NASA's hope is that the space station will help draw new talent to the field. Those contemplating entering microgravity research may have to wait several years until the station is fully assembled. But what ground-based scientist would expect to start work in a new laboratory while builders were still installing the plumbing? Space is no different, just more complicated.

Scientists who have in the past been sceptical about space-based research should recognize and applaud the changes under way at NASA. The station has never been exclusively a science project, and should not be judged as such. It is most impressive as a feat of off-planet engineering, and it exists primarily because the United States and its partners want to establish a continuous human presence in space. For scientists, however, Alpha offers a real chance, at last, to find out whether there are substantive research questions worth pursuing on the high frontier. ■

Biotechnology battleground

An independent inquiry is needed to restore morale at an international biotechnology research centre in New Delhi.

Ever since the potential of genetic engineering became apparent, it has been clear that a major beneficiary, in terms of its applications to both human health and enhanced food production, could be the developing world. There was therefore much enthusiasm in the early 1980s for the proposal, spearheaded by the United Nations Industrial Development Organization (UNIDO), to set up an international centre dedicated to help achieve this. The International Centre for Genetic Engineering and Biotechnology (ICGEB), as it became known, was accordingly set up with twin 'components' — one in Trieste, Italy, and the other in New Delhi, India.

In its 13 years of existence, the centre appears to have achieved much to its credit. Thousands have taken its training courses, and researchers at the two components have a growing list of publications in international journals (see <http://www.icgeb.trieste.it>). But these successes appear in danger of being undermined by bitter disputes in recent months about the operation of the New Delhi component (see page 127). The disputes centre around allegations of lax management, external pressure on appointments, and fears of intimidation among staff. The result, at least according to those making the allegations, is a loss of morale among researchers and a rapid turnover of junior members, particularly PhD students.

The directors of both the ICGEB itself and the New Delhi compo-

nent insist that, although mistakes may have been made, the broader allegations are ill-founded and essentially "mischievous". Certainly it is impossible, without knowing the full details of each case, to make a definitive judgement on the overall situation. The ICGEB had a difficult gestation, marred by a sitting dispute that was even more bitter than usual. Apart from Italy and Russia, no large industrialized nation has joined, and the resulting financial pressures have been worsened by the slowness of some of the other 41 members to pay their dues.

Such pressures have inevitably created their own tensions. But the strength of the allegations about the working environment at the New Delhi centre, and the possibility that the scientific vitality of the whole exercise could be undermined, need to be treated seriously. It is essential, for example, that senior appointments in an institution seeking a high international profile are made in a way that engenders trust, not suspicion — and that staffing adequately reflects the international diversity of the goals and activities of the centre itself.

There have been calls for an independent inquiry into the operation of the New Delhi component. Such a commitment should not be undertaken lightly, if only for financial reasons. But the proposal should be considered seriously by ICGEB board members when they meet in New Delhi next week. Internal tensions are clearly running so high that perhaps only an objective, outside view can ascertain the truth. ■





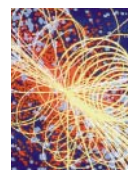
Chemical caution
Study suggests link
between pesticides
and Parkinson's
p125



Weather eye
Climatologist made
president of French
research agency
p126



Trouble in Delhi
Indian genetics
centre beset by
staff criticism
p127



Higgs dilemma
Particle physics lab
faces a \$56 million
question
p128

Emphasis of NASA's microgravity research shifts to space biology

Tony Reichhardt, Washington

As the long-awaited International Space Station becomes a reality — the first crew of one astronaut and two Russian cosmonauts boarded last week — the US space agency NASA is redirecting its controversial microgravity research programme towards the life sciences.

Future microgravity research will focus more on biology and biologically inspired technology, and less on the kinds of crystal-growth experiments that yielded questionable results in the space-shuttle era.

Under the leadership of chief scientist Kathie Olsen, the agency has already reorganized its Office of Life and Microgravity Sciences and Applications, and given it a new name, the Office of Biological and Physical Research (OBPR).

More importantly, the office has been made an 'enterprise', on a par with human spaceflight and space science. The move reflects the importance that administrator Daniel Goldin has attached to beefing up his agency's capability in the biological sciences.

Nobel prize-winner Baruch Blumberg, who heads the agency's Astrobiology Institute, has been named a senior adviser to Goldin to help achieve this goal. Olsen will be acting administrator of the renamed office until a permanent leader can be found.

The creation of the OBPR is the latest of several shifts in NASA's microgravity and biology research programme in recent months. Long-time head of life sciences at NASA, Arnauld Nicogossian, who is widely



Action station: Russia's mission control welcomes the first crew to the International Space Station.

seen as more interested in astronaut health and 'space medicine' than other aspects of space biology, has been reassigned as the chief medical and health officer.

Other members of the old guard in the programme have retired, creating opportunities to hire younger scientists more versed in molecular biology and genomics.

Olsen and Blumberg are seeking leaders for OBPR's five restructured divisions: physical sciences (which incorporates much of the old microgravity research portfolio along with some new thrusts in nanotechnology); fundamental space biology; biomedical and human support research, which focuses on keeping astronauts healthy; research integration; and policy and programme integration.

The biggest changes will be in the physical

sciences division, which spends about 40% of the office's \$317 million annual budget. Before, 90% of this went on basic research into the behaviour of materials and processes such as combustion in microgravity. Now, half of the research will be more applied, says the division's director, Eugene Trinh.

Work in different disciplines may be consolidated and streamlined, says Trinh. He adds that it will also be more focused on NASA's need to assure the safety of astronauts and their spacecraft.

The physical sciences division will also oversee a technology initiative dear to Goldin's heart — the development of tools and techniques in biomolecular physics that could someday lead to 'nanobots' capable of coursing through a Mars-bound astronaut's body, diagnosing and repairing disease without the need for a doctor.

NASA and the National Cancer Institute have formed a partnership to explore such research. Goldin envisages each agency spending around \$10 million a year for five years, although the cancer institute will invest just \$6 million this year. NASA will match that, and add a further \$4 million for related nanotechnology work.

Another key shift in the microgravity research programme is a retreat from NASA's long-hyped emphasis on protein crystal growth as a boon to drug developers and biotechnologists.

US grad students win union rights

Paul Smaglik, Washington

US graduate students working as research and teaching assistants at private colleges and universities have the right to form unions, the National Labor Relations Board ruled last week.

The case, which involves graduate student assistants at New York University, reverses a position the board has taken since

the mid-1970s. The ruling could speed up the unionization of graduate students in private universities across the country.

Unionization at public universities has begun in several states, including Wisconsin, Michigan, California and New York. Attempts to unionize by graduate assistants in public universities are covered under state law, not under the National Labor Relations Act.

A critical National Research Council panel found earlier this year that “one cannot point to a single case where a space-based crystallization effort was the crucial step in achieving a landmark scientific result” (see *Nature* **404**, 114; 2000). The panel challenged NASA to instigate a grant programme to identify macromolecules important to structural biologists, which could be used to test the value of space-grown crystals once and for all.

NASA issued just such a call for research, with proposals due two weeks ago. Gary Stein of the University of Massachusetts Medical School, who chaired the panel at the time of the report, and who does not himself conduct space research, says: “I’ve never seen a report given to any group, let alone a government agency, where the recommendations were followed to this extent and with this speed.”

Congress recently added \$25 million to NASA’s 2001 budget for ground-based microgravity research, and ordered the agency to come up with a plan for doing research on the space shuttle while the space station is being built.

Opportunities for small-scale experiments are being explored for life scientists, says Olsen, including experiments ‘piggy-backed’ on other satellites.

But the space station will provide the most important opportunities. A crew that returned last week sailed through its construction tasks with apparent ease. Another crew will attach the US laboratory module in January.

At first, construction will take precedence over science. “I don’t think in the immediate future we can have incredibly high expectations,” says Stein.

But as the station nears completion in 2005, better laboratory equipment, along with the chance to conduct long-term studies and repeat experiments, should create a “whole new generation of opportunities” in space research, he says. ■



Gently does it: ground-based technicians assemble the space station’s laboratory.

Mexican science waits to see if election promises are met

Marina Chicurel, Santa Cruz

Plans drawn up by president-elect Vicente Fox to bolster science and technology have received a cautious welcome from Mexican scientists. Scheduled to assume power on 1 December, Fox is the first president in 70 years to represent a political party other than the Institutional Revolutionary Party.

Fox gained a reputation as a supporter of science when he was governor of the state of Guanajuato from 1995 to 1999. And he has promised to double research and development (R&D) spending as a proportion of Mexico’s gross domestic product, and to increase the number of scientists.

Although they welcome his plans for science, many researchers are sceptical about how much will be delivered, given the current shortage of funding, the track record of past governments, and the lack of detail in the commitments.

But the team’s coordinator of science and technology, María del Carmen Díaz, says big changes are planned. In addition to extra funding, the team promises to improve links between the academic world and industry, and to boost the political standing of science.

“Science seems to be more present in the minds of this transition cabinet than before,” says neuroscientist René Drucker Colín, president of Mexico’s National Academy of Sciences. “But science is not a political priority.”

Mexican scientists have long complained about their lack of political status (see *Nature* **398**, suppl. A7–A9; 1999). The country’s main research funding agency, CONACYT, an arm of the Ministry of Public Education, has to compete with education for its budget.

To improve the situation, the academies of science, engineering and medicine recently proposed that CONACYT become autonomous, receiving its budget from Congress and answering directly to the president.

“Such a change in organization would send a clear message that science and technology are as important as human rights or clean elections,” says Miguel José Yacamán, a physicist at the National Autonomous University of Mexico (UNAM) in Mexico City and former associate director of CONACYT, referring to the recent successes of other autonomous organizations, such as the National Commission for Human Rights and the Federal Electoral Institute.

Fox’s team wants to relax CONACYT’s link to education. Tessy López, a chemist at the Metropolitan Autonomous University in Iztapalapa, is an unofficial scientific adviser to Fox’s team. She favours the less radical move of splitting education between two ministries: one to focus on postgraduate education and



Fox: has a record of supporting science.

research, and the other covering all other levels of education.

Díaz thinks this is not necessarily the best option, and Yacamán thinks it would not go far enough. But López says it would allow CONACYT to increase its power more gradually, and thus more effectively.

Another concern is the poor link between the academic world and industry. Economist Axel Didriksson of UNAM points out that the percentage of

research and development financed by private industry in Mexico is less than a third of that of developed countries.

Many scientists are sceptical about whether the situation will improve, however, given that most of the international companies that have business interests in Mexico do their research in their countries of origin. And many Mexican companies are reluctant to invest in risky, long-term projects.

But researchers who dealt with Fox when he was governor of Guanajuato are optimistic. Mechanical engineer Arturo Lara, director of the Council for Science and Technology for the state, says that Fox established the first direct links between the state government, researchers and local industry, when he created the council in 1996.

As an example of the council’s achievements, Lara cites an increase of about 30% in the research and development budget of the state’s main research institution, the University of Guanajuato.

Octavio Paredes, a biotechnologist at the Centre for Research and Advanced Studies of the National Polytechnic Institute (CINVESTAV) in Irapuato, says that, in contrast to previous governors, Fox listened to scientists’ concerns. He took an active role in trying to deal with them, visiting research laboratories and inviting scientists to his office.

Fox also tried to implement long-term science policies, an innovation that would be widely welcomed at the national level. Paredes says the transition team is developing plans spanning up to 30 years, including an increase in the number of research centres in the country.

But all these plans are just words unless the new government manages to funnel more money into science. ■

Pesticide use linked to Parkinson's disease

David Adam

Long-term low-level exposure to a pesticide widely used in horticulture and water management could be a cause of Parkinson's disease in humans, researchers at Emory University in Georgia warned this week.

They found that rats given repeated doses of the insecticide rotenone, which is derived from plant extracts, have difficulty walking, shaky paws, and protein deposits in their brains — all tell-tale symptoms of Parkinson's disease.

The results, which appear to strengthen existing epidemiological evidence of a link between pesticide exposure and Parkinson's, were announced at a meeting of the Society for Neuroscience in New Orleans. "We've shown that exposure is sufficient to do it in rats, and presumably the same can happen in people," says neurologist Tim Greenamyre, a member of the Emory team.

Parkinson's disease affects one or two adults in every thousand, and the risk increases tenfold beyond the age of 50. Sufferers become clumsy and move stiffly, and their muscles often spasm. These and other

symptoms occur because neurons that produce dopamine, a chemical messenger that rouses the body to action, degrade. In addition, fibrous protein deposits called Lewy bodies accumulate in brain cells.

Rotenone is extracted from several plants whose juice has been used by South and Central American people for centuries to catch fish. Unlike synthetic pesticides, it breaks down quickly in the environment, and was thus thought to pose little danger.

To study the physiological effects of rotenone, the pesticide was injected into the bloodstream of laboratory rats. Like many other pesticides, it was known to block the activity of the enzyme needed to provide the dopamine neurons with energy. In the rats, rotenone killed some dopamine neurons and caused deposits resembling Lewy bodies.

Benoit Giasson, a Parkinson's disease specialist at the University of Pennsylvania, Philadelphia, points out that the doses given to the rats — and their administration through injection — are very different from the normal human exposure. But he still

argues that rotenone should be used with greater caution until more is known about the risks.

Because rats injected with rotenone mirror the symptoms of Parkinson's so closely, the animals will help researchers to study the disease. The work will be published in the December issue of *Nature Neuroscience*. ■



Unprecedented gift boosts basic physics in Canada

David Spurgeon, Montreal

An independent institute for research into theoretical physics is being set up in Waterloo, Ontario, thanks to the generosity of a Greek immigrant who made his fortune through creating a high-tech firm that specializes in hand-held electronic communication devices.

The Perimeter Institute for Theoretical Physics is being financed with a donation of Can\$100 million (US\$65 million) from Mike Lazaridis, the 37-year-old founder and joint chief executive officer of Research In Motion.

It is believed to be the largest philanthropic gift in Canadian history and the country's biggest single private contribution in support of basic science.

The institute will be located on a site that will be provided at a nominal rent by the city of Waterloo. The gift from Lazaridis is being supplemented by a further Can\$10 million each from the company's other joint chief executive officer, Jim Balsillie, and its vice-president of operations, Doug Pregon. The initial donations are hoped eventually to become a permanent endowment.

The proposed institute is planned to have links with all of the universities in the region, and will initially operate with a core of between 10 and 15 researchers, who may have joint appointments with other institutions. Areas of research to be covered by the institute include quantum gravity (both string and non-string forms), quantum information theory and quantum computing, elementary particles and fields, and cosmology.

Lazaridis, who arrived in Windsor, Ontario, with his parents in 1967, graduated from the University of Waterloo in electrical engineering and founded Research In Motion in 1984. The company describes itself as "a world leader in the mobile communications market", and had sales last

year of US\$120 million.

Lazaridis says that basic science in Canada is underfunded, and that the gifts will support an area of research that is vital to the technological advances that benefit humankind.

"Just about every major industrial revolution and technical advancement is [rooted] in basic theoretical physics research," he told a press conference announcing the creation of the institute.

He cited research on lasers, wireless technology, computers, semiconductors and superconductors. "It all started with thinkers thinking about problems and laws that are the foundation of our physical world," he said.

The creation of the institute has been described by John Polanyi of the University of Toronto, who won the Nobel prize in chemistry in 1986 for his work on elementary chemical processes, as "a gesture of historic importance for Canadian science". Polanyi described the gift as "entirely out of the mainstream".

Polanyi also called for government to fall into line with its science funding. The federal government of Canada and that of Ontario province are already negotiating with Perimeter officials about their possible involvement, according to its executive director, Howard Burton. ■



Basic instinct: entrepreneur Mike Lazaridis unveils the plans for the Perimeter Institute.

Court young French postdocs, says petition...

Declan Butler, Paris

High on the agenda for Gérard Mégie (see right) and Geneviève Berger, the two new leaders at France's Centre National de la Recherche Scientifique (CNRS), is the need to respond to many French scientists' worries about the future of the country's research base.

Last week, members of the Conseil Supérieur de la Recherche et de la Technologie, the research ministry's main advisory body, relaunched a public petition calling for a long-term scientific employment policy.

The move follows a recent warning by the Academy of Sciences that many research areas — particularly genomics, nuclear science and pharmacology — face a crisis if more is not done to attract new students and to retain postdocs. The age profile of researchers in France means that some 50% will need replacing over the next ten years.

Henri-Edouard Audier, head of a chemistry laboratory at the Ecole Polytechnique near Paris and trade union representative on the Conseil Supérieur, says that, although the 2001 research budget is "the best in ten years", more must be done to secure a long-term commitment to young researchers. This is needed, he says, to attract more young people to a career in research and stem the postdoc brain drain.

The 2001 budget (see *Nature* 407, 435–436; 2000) provides for 265 new research posts, and research minister Roger-Gérard Schwartzberg has said that he is committed to a long-term employment policy for science. But many of the posts falling empty are teaching posts in universities, which come under the ministry of education, rather than the research ministry.

The petition's signatories, who include nearly 1,400 eminent French scientists, want reassurance that any employment policy will be a collaboration between the two ministries. Audier says he hopes that Mégie, who signed the petition in his previous job as a university professor, will remain "consistent" in his CNRS role.

But Olivier Laprévotte, a biochemist who runs a laboratory on the CNRS campus at Gif-sur-Yvette, believes that government action comes too late. He says there are already laboratories without directors, and too few young researchers with the experience to replace them within the next few years.

The problem goes beyond the lack of an employment policy. Laprévotte asks: "Even if we can promise long-term career opportunities to doctoral students, who will want to take them up?"

"Newly qualified researchers, even with

...as changes at the top suggest action ahead

The French government announced last week the appointment of climatologist Gérard Mégie as president of the country's main research agency, the Centre National de la Recherche Scientifique (CNRS).

Research minister Roger-Gérard Schwartzberg says that the choice of Mégie, currently director of the Institut Pierre-Simon Laplace des Sciences de l'Environnement Global, based in Paris and Versailles, reflects the government's desire "to bring science and environment closer together".

It also reflects a drive to develop multidisciplinary at the CNRS. Mégie, a physicist turned environmental scientist who specializes in atmospheric research, joins Geneviève Berger, a physicist turned biologist, who was named CNRS director-general in August (see *Nature* 407, 435–436; 2000).

Schwartzberg says the pair will give "a new boost of life to a CNRS that is keeping up with the



Mégie: 'will implement recently approved reforms'.

times". They will implement reforms to the organization approved two weeks ago by the Conseil d'Etat, France's top legislative body, after a lengthy national debate (see *Nature* 404, 426; 2000).

These reforms give the agency greater independence from the research ministry and a clearer division of responsibilities at the top. The president of

the agency, together with its administrative council, will focus on overall policy and manage the CNRS's relationship with external partners, including universities and international organizations.

In contrast, the director-general will be responsible mainly for the scientific, administrative and financial management of the agency, and will have the authority to create new scientific departments and institutes without waiting for the research ministry's approval.

A new external evaluation committee will be formed, made up of French and foreign scientists and industry representatives, to evaluate the activities of the CNRS at least every four years. The main scientific and departmental advisory councils will be broadened to include more foreign members, and a new ethics committee will be created.

D.B.

two or three years of postdoc experience, are forced to go through a complicated and, at times, random recruitment process before clinching a post that will offer them a starting salary of at best between 10,000 and 11,000

francs (US\$1,300–1,450) per month, a fraction of what they can earn in the private sector. It is difficult to find an argument to persuade young researchers to stay in the public sector." ■

Japan's GM corn will undergo tests

David Cyranoski, Tokyo

Reports that an unapproved form of genetically modified corn (maize) has found its way into the Japanese food supply have prompted government action in both Japan and the United States.

StarLink corn, produced in the United States by the French company Aventis CropScience, contains the pesticide Cry9C, which some have warned could cause allergic reactions if consumed. As a result, it has been approved for use only in animal feed in the United States, and the discovery of Cry9C in taco shells on sale in US supermarkets led to a massive product recall earlier this year (see *Nature* 407, 438; 2000).

In Japan, StarLink has not yet been

approved for consumption by either humans or animals. But the Consumers Union of Japan reported late last month that StarLink protein had been detected in corn meal, and there have been reports of it showing up in six other products.

The US Department of Agriculture has now announced that it will test corn that is being shipped to Japan for the presence of StarLink, and Japan's equivalent ministry will start doing its own tests on the product's safety. These tests include feeding the corn to chickens and testing for any debilitating effects on the chickens themselves, and for the presence of the Cry9C protein in the chicken meat. ■

Indian biotech centre rocked by controversy

K. S. Jayaraman, New Delhi

When the board of governors for the International Centre for Genetic Engineering and Biotechnology (ICGEB) meets next week, it may be in for a shock. The board, which represents the ICGEB's 43 member states, will find the centre's New Delhi component embroiled in controversy over allegations of insufficient funding, staff harassment and political influence on the recruitment of researchers and officials.

Set up 13 years ago by the United Nations Industrial Development Organization (UNIDO), the ICGEB was intended to spearhead the transfer of western biotechnology to developing nations. The centre is split into two components, one based in New Delhi, India, and the other in Trieste, Italy.

Earlier this year, officials at the New Delhi component denied reports that budget cuts and deteriorating working conditions had caused many scientists to leave (see *Nature* 403, 694 and 404, 329; 2000). But the complaints continue.

For example, the son of a senior official in India's Department of Biotechnology — part of the Ministry of Science and Technology — has been appointed as an international fellow at the ICGEB. The move has raised a few eyebrows, as such positions are intended to be reserved for scientists from member countries other than India.

And Honey Reddi, a researcher in the virology department, claims she was "harassed" into resigning and then barred from entering the laboratory after she complained openly about working conditions and recruitment practices at the centre. She has now lodged formal complaints with India's National Commission for Women and its National Human Rights Commission.

There have also been accusations of lax management procedures. S. K. Panda, for example, a professor of pathology at the All India Institute of Medical Sciences in New Delhi and a former consultant to the centre, holds a patent on the full-length clone of the



Barred? The ICGEB faces claims from Honey Reddi (inset) that she was "harassed" into resigning.

hepatitis E virus genome. He says that the ICGEB sent the clone to laboratories outside India without his knowledge, violating the material transfer agreement he had signed with the centre.

Although Panda's claim is currently under investigation, Arturo Falaschi, the Italian director of the ICGEB, dismisses the other complaints as "trivial". Virander Singh Chauhan, head of the Delhi component, describes allegations of political appointments as no more than "mischievous" propaganda.

Yashwantrao Vaishnav, who was a senior research scientist in the virology group until resigning in 1998, complains that "the politics is so deep-rooted at ICGEB that, it appears to me, no significant change can be brought about".

Vaishnav, who now works outside India after being one of 17 scientists to quit during the past two years, says that, although talented researchers remain, they are not being used properly because "the atmosphere at ICGEB is not conducive to it".

Chauhan denies there is anything wrong at the centre. "For the past two years, we have been continuously hiring people," he says. "Most of the people who left were here either on project positions or on short-term fellowships."

Responding to charges that the appointments process lacks transparency, Chauhan says: "On the contrary, I think we are too open ... not even a technician is recruited without advertising for the post and a selection committee."

But Devaguptapu Subrahmanyam, who was closely associated with the development of the ICGEB as a UNIDO official and later as a senior adviser at the New Delhi component, argues that the centre's recruitment procedures "need to be revamped". In partic-

ular, he says, all senior positions "should be filled by advertisement through international media".

"There are clearly some serious problems with this unit, and an external review is needed to get to the bottom of it," says Deshpal Verma, professor of molecular genetics at Ohio State University. Verma was first choice to head the New Delhi unit when it was set up by UNIDO in 1987, but rejected the offer.

Like Chauhan, several members of the ICGEB's Council of Scientific Advisors (CSA) reject criticism of its operation. "The scientific contributions of the Trieste and New Delhi laboratories are world class, and the training and help given the centre's member states are of extraordinary value," says one, Nobel laureate Arthur Kornberg of Stanford University School of Medicine.

Another CSA member, Govindarajan Padmanaban of the Indian Institute of Science in Bangalore, says he is not aware of internal problems at the centre. "My overall impression of ICGEB is positive," he says.

Contributing to the current tension is concern over future funding prospects for the centre. According to Subrahmanyam, its long-term financial stability "looks uncertain", as a number of member countries are behind with their dues, and the two host countries — India and Italy — are both showing signs of "donor fatigue".

Subrahmanyam argues that it is now vital for the ICGEB to persuade other developed countries to become members. But this will mean giving the centre a stronger international profile. At present, he points out, hardly any of the senior positions in the New Delhi component are held by non-Indians.

♦ <http://www.icgeb.trieste.it>



Panda: in dispute with the ICGEB over its handling of his clone of the hepatitis E genome.

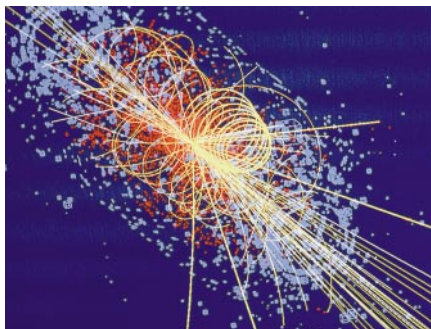
CERN agonizes over whether to keep up the hunt for Higgs

Munich Senior managers at CERN, the European Laboratory for Particle Physics, were meeting as *Nature* went to press to decide whether to fund a one-year extension for the Large Electron–Positron (LEP) collider, at an estimated cost of 100 million Swiss francs (US\$57 million), in a bid to confirm earlier possible sightings of the elusive Higgs boson.

The LEP was switched off this week after a one-month extension to its planned running time. This initial extension was based on hints that the particle had been seen in two of the collider's detectors during the summer (see *Nature* **407**, 118 and **408**, 10; 2000).

A further possible sighting means that the chances that Higgs has not been seen are four in one thousand. But high-energy physicists usually accept only a chance of one in ten million as conclusive, as the formal uncertainty in an experimental result is often an overestimate of the true robustness.

For this to be reached the LEP would have to keep running for one more year, and would have to be upgraded to run at a slightly higher energy. This could be done



Is it? Experiments at CERN are still showing hints of the decay of a Higgs boson (here simulated).

easily, says Tiziano Camporesi, head of one of the LEP experiments, using four superconducting magnets previously used as test or demonstration models.

Physics Nobel backs cash bid for Russian electronics

Moscow Zhores Alferov, recent Nobel prize-winner for physics and a communist member of the lower chamber of the Russian parliament, last week spoke in support of the nation's electronic industry. At a parliamentary session on the prospects for Russian electronics, he said that the collapse of the Soviet Union had damaged this segment of the economy "more than

anything else", and that it now meets only 15% of the country's needs.

The parliament voted an extra 440 million rubles (US\$16 million) into the 2001 budget for a programme to develop electronic technologies. If the cabinet approves the bill, it is expected to increase electronics production fivefold by the end of 2006.

Latin America signs up to carbon trading

Washington Fourteen Latin American countries have agreed to back a controversial bid by the United States to allow it to 'buy' greenhouse-gas emission reductions and the use of carbon sinks (the planting or saving of forests) from poorer nations. The deal, announced shortly before the latest round of international negotiations on climate change due to open in The Hague next week, will help the United States meet its reduction quota agreed at the previous summit in Kyoto, Japan, in 1997.

The Kyoto Protocol allows developed countries to fund projects in developing countries, but does not make it clear if these count as reductions. Europe opposes such emissions trading and the use of carbon sinks as incentives, because the sinks are not permanent and can release carbon dioxide when burnt.

New European computer network takes off

Brussels European efforts to construct a computing infrastructure for the twenty-first century were officially launched last week. A consortium of 30 national research networks has signed an 80 million euro (US\$73 million) contract with the European Commission to develop the Géant research Internet network.

Géant, which will link more than 3,000 institutions, will replace and expand the existing European TEN-155 network (see *Nature* 405, 261–262). It will boost the network's capacity from about 155 megabits per second to 2.5 gigabits per second, rising to 10 gigabits per second by 2005. Géant will also lay the foundations for an advanced research computing grid.

US librarians fight publishing merger

Washington Officials from the Association of Research Libraries (ARL) told attorneys from the anti-trust division of the US Department of Justice on Monday that a merger between the giant British–Dutch publisher, Reed Elsevier, and Harcourt General, a US rival based in Chestnut Hill, Massachusetts, would inflate prices and

reduce access to hundreds of science, technology and medical journals.

But John Regazzi, president and chief executive officer of Elsevier Science, Reed Elsevier's North American science publisher, says that the company has held percentage price increases to single figures in the past two years. He says that this policy will continue for the newly acquired Harcourt journals. The justice department must, along with the European Union, approve the \$4.5 billion merger, which was announced on 27 October.

UK government reviews animal experiments policy

London British biomedical scientists, frustrated by months of delays in approving research on animals, may soon get some relief. The Home Office, which regulates animal experiments, says it will review the operation of a local ethical review process, introduced in April 1999 and blamed for the delays.

The move is not, in itself, a surprise — an examination of the measure's effectiveness was promised before it was introduced. But scientists are heartened by the statement that the review, to be completed by the middle of next year, will consider "what problems may have been encountered".

Californian university gets money for women

San Diego An anonymous donor has made a grant of \$20 million to increase the number of women faculty members in science and engineering at the University of Southern California (USC), Los Angeles. The grant will fund new faculty positions for women, upgrade laboratories, increase scholarship aid for undergraduates, create new fellowships for graduates and fund child care. USC Provost Lloyd Armstrong Jr described the money as a "very exciting and visionary gift".

USGS geologist to head ocean drilling programme

Washington Steven Bohlen, currently an associate chief geologist at the US Geological Survey, has been named president of the Joint Oceanographic Institutions (JOI) and executive director of the Ocean Drilling Program (ODP).

JOI and the US Science Support Program have developed a CD-ROM virtual tour of the ODP's drill ship *JOIDES Resolution*. Using sediment cores recovered by the ship, students can study the dynamics of ice sheets over the past four million years.

► <http://www.joi-odp.org>

Structures by numbers

Advances in automation and genome sequence data will allow new protein structures to be produced faster than ever before. As the era of structural genomics unfolds, it could revolutionize drug discovery, says Alison Abbott.

It took 22 years of painstaking effort for Max Perutz to produce his structure for the protein haemoglobin¹. Published in *Nature* in 1960, the purified molecule's precise configuration was calculated from the way its crystals diffracted a beam of X-rays. Since this pioneering work, the number of protein structures that have been determined has increased exponentially. So much so that the field's main database, the Protein Data Bank, now contains some 13,000 unique structures into which proteins can fold (see chart, opposite).

But such rapid progress will soon appear sluggish. The rate at which protein structures are cranked out is poised to rise by one, or even two, orders of magnitude. As the DNA sequencing stages of the various genome pro-

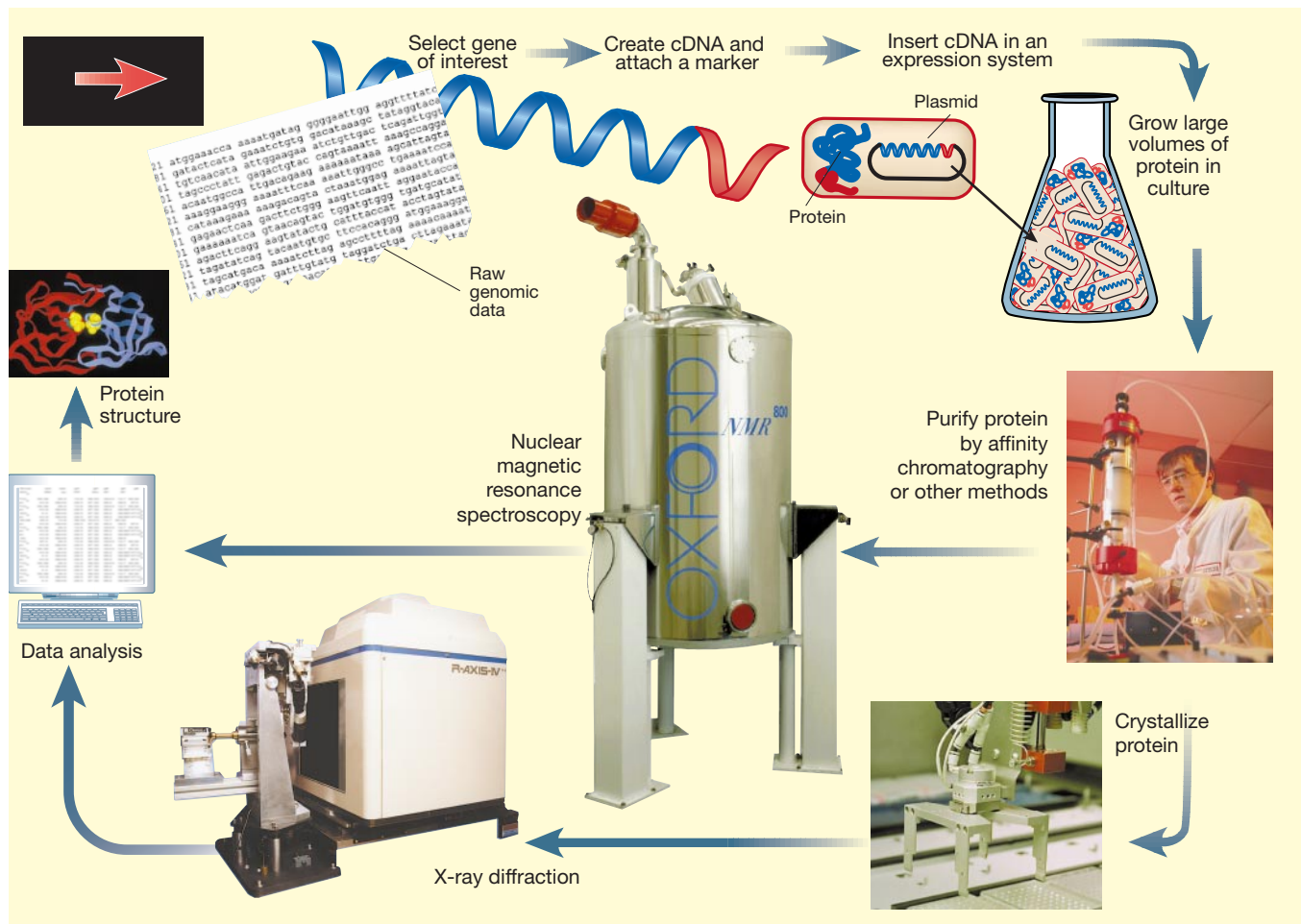
jects come to a close, attention is turning to the functions of the proteins encoded by the genes. The same brute-force philosophy that has turned genetics into genomics is now beginning to transform structural biology into structural genomics. "Structure and sequence information together are more insightful than sequence information alone," says Gaetano Montelione, a structural biologist at Rutgers University in New Jersey.

This new discipline — which is the focus of a special *Nature Structural Biology* supplement this month² — promises to change fundamentally the way structural biologists work. Getting basic structural information will largely be automated as huge 'protein structure factories' transform raw genomic data into finished protein structures on an

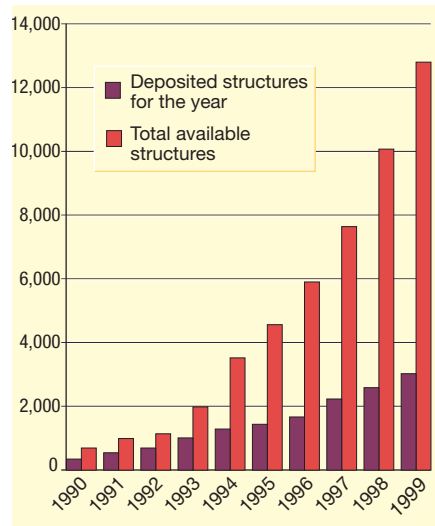
unprecedented scale. This will leave academic structural biologists free to concentrate on interpreting the biological meaning of the new information, while those in the pharmaceutical industry apply the resulting knowledge to drug design.

On the production line

A structural genomics production line (see below) starts with a gene. Interest focuses on the parts of the gene that are transcribed into messenger RNA, and so encode proteins. Synthesizing sequences of DNA — known as complementary DNA (cDNA) — that correspond to these sections paves the way for generating the relevant protein on a large scale. By adding short pieces of marker DNA to the cDNAs, the final protein gets



On the conveyor belt: structural genomics will turn protein structure discovery into a factory production-line process.



Data deluge: the number of structures in the Protein Data Bank has grown steadily.

'tagged' with a particular peptide, which makes extraction of the product easier.

Each cDNA is then transferred to a suitable expression system. At present, this usually involves splicing the cDNA into a loop of DNA known as a plasmid, which in turn is inserted into *Escherichia coli* bacteria. Culturing the bacteria produces the protein. To purify the product, the homogenized culture is normally poured through an affinity chromatography column containing molecules that bind specifically to the peptide tag. For example, proteins tagged with a series of histidine residues bind tightly to nickel ions.

The protein is then washed from the column and its structure determined using two main techniques: X-ray diffraction or nuclear magnetic resonance (NMR) spectroscopy. Each has its advantages, and most protein structure factories will use both methods. Because it works on molecules in solution, NMR is used to study the dynamics of proteins 'in action', for example recording the structural changes that occur when a protein binds to a small molecule — such as a hormone or a drug.

At the moment, NMR cannot be used with molecules larger than about 30 kilodaltons, and so cannot study large proteins in their entirety. But the technique is good enough for studying individual 'domain folds', twists and turns within distinct areas of any protein. There may be fewer than 20,000 possible configurations of these folds.

X-ray diffraction is not constrained by the size of the protein, and has the added advantage that producing a structure from the raw data requires much less computation than for NMR. But it does depend on being able to crystallize the protein — and determining the precise conditions to achieve this can be very laborious.

The great potential of structural genomics has already been demonstrated. In



Heinemann: plans to focus on amenable human proteins with therapeutic potential.

the past couple of years, around 70 proteins have had their structures determined using DNA sequence data as the starting material. All that remains is for the 'factory' element to be established. And structural genomics centres worldwide are making progress towards automating the entire production line.

Taking the initiative

Surprisingly, Germany was among the first to take the leap. Research ministry cash started flowing into the Protein Structure Factory, at the Max Delbrück Centre for Molecular Medicine in Berlin, more than a year ago. "Usually, German agencies decline to invest in big projects which have not already 'proven their importance' by being supported in the United States first," says Udo Heinemann, scientific spokesman for the initiative.

This time, the US National Institutes of Health (NIH) followed Germany's lead. In September it funded seven centres to develop technologies for high-throughput structural genomics. Elsewhere, the RIKEN Structural Genomics Initiative in Yokohama, Japan, last month switched on its battery of 16 NMR machines; the Ontario Center for Structural Proteomics was launched in Toronto in July; and other initiatives are getting under way in France, Switzerland and Italy.

Automating the protein structure pro-

duction line is reasonably straightforward — even the fiddly step of crystallization can be handed over to computer-controlled robots, which vary subtly the conditions in hundreds of thousands of wells containing protein solutions. The problem is that many proteins cannot readily be expressed in culture, either failing to fold correctly or precipitating out into messy clumps, rather than remaining in solution. Proteins that are bound within cell membranes are particularly troublesome, as they do not fold correctly in aqueous solution. And even those proteins that are soluble can sometimes be difficult to crystallize.

Given these problems, some of the structural genomics initiatives have decided to concentrate on the 'low-hanging fruit' of amenable proteins first. Guido Granti of Chiron Vaccines in Siena, who is coordinating an Italian structural genomics consortium now being set up, hopes to identify new target proteins suitable for vaccine development, particularly against meningococcus B, which causes meningitis. The consortium will concentrate on the 25% of proteins from cloned bacteria that Chiron has managed to express easily. "It means abandoning many proteins that might have been interesting," says Granti. "But we will have plenty to be working on while deciding if we want to attack insoluble proteins."

In Berlin, the focus is on human proteins, which are notoriously difficult to study because they are extensively modified after being translated from messenger RNA. "We select our target proteins on the basis of their potential novelty, and their potential for involvement in disease," says Heinemann. But, like the Italian effort, the initial focus is on proteins that are, as Heinemann puts it, "nice to us".

Many researchers are experimenting with a number of tricks to coax greater coopera-

It will be at least five years before structural genomics has an impact on biology — then it will explode.

tion from intractable proteins. This includes randomly mutating the gene in question to see if any one of numerous point mutations produces a more soluble protein. For example, Tom Terwilliger of the Los Alamos National Laboratory in New Mexico, who is coordinating one of the NIH-funded structural genomics centres, is using this approach in his assault on 1,000 of the 4,000 proteins produced by the tuberculosis bacterium.

Alternative techniques include finding molecules that will bind to a target protein, stabilizing it in its correct structure, or expressing the protein of interest alongside 'chaperone' proteins which help direct its folding. Shigeyuki Yokoyama, leader of the Japanese initiative, plans to use a cell-free expression system incorporating a 'universal chaperone' protein. Initially, the Japanese effort will focus on mouse proteins, as the RIKEN Genomic Sciences Center, which last week hosted the International Conference on Structural Genomics 2000, has already sequenced thousands of full-length mouse cDNAs.

Other researchers are targeting entire protein families — the difficult molecules as well as the easy. Dino Moras, of the Institute of Genetics, Molecular and Cellular Biology in Strasbourg, is coordinator of the French structural genomics initiative. As a whole, the initiative has a broad outlook, but Moras's own group will concentrate on receptors found in the cell nucleus — proteins such as the oestrogen and glucocorticoid receptors — which are already the focus for work in his lab.

Yet others, such as the planned Swiss consortium, coordinated by Markus Grütter of the University of Zurich and Kurt Wüthrich of the Swiss Federal Institute of Technology, also in Zurich, have taken the bold decision to specialize in difficult proteins. The Swiss effort will pay particular attention to membrane proteins, which account for at least half of all cellular proteins.

On target

In general, the targets that each initiative has selected reflect the research interests of the scientists involved. Initially, it was suggested that proteins might be distributed across the various structural genomics initiatives just as responsibility for sequencing chromosomes was divided up among the teams working on the international Human Genome Project. "But the idea that targets could be imposed on the community evaporated when it was realized that there was never going to be any agreement on an ideal target," says Stephen Burley of the Rockefeller University in New York, who directs one of the NIH structural genomics centres.

The resulting lack of an overall strategy for the field is one reason why the pharmaceutical industry is still largely watching from the sidelines (see 'Drug companies bid their

Drug companies bid their time

'Rational drug design' has been the catch-phrase among medicinal chemists since the 1980s. Although, at first, little came out of the attempts to use three-dimensional protein structures to help design the perfect drug, the past few years have seen the approach notch up some notable successes. Recent designer classics include Glaxo Wellcome's anti-influenza drug Relenza³ and Merck's HIV treatment, Crixivan⁴.

Pharmaceutical companies are very interested in having access to the tens of thousands of protein structures that structural genomics could deliver. And at the First International Structural Genomics Meeting in April this year, scientists in industry and academia decided to investigate the possibility of creating a structural genomics consortium to speed up international efforts.

The plan is to use the SNP (single-nucleotide polymorphism) consortium as a model. Made up of 13 companies and the Wellcome Trust, the SNP consortium supports the identification of the most common mutations in the human genome. It was organized by Alan Williamson, a former vice-president of

research at Merck, who is also putting the structural genomics consortium together. But his new task is not so straightforward. "When we set up the SNP consortium, the technology — high-throughput sequencers — was already in place," says Williamson. "Potential partners could go and see what was going on in public laboratories and judge how soon the goods could be delivered." Meaning also, of course, that they knew how much it would cost.

Structural genomics is at an earlier stage of development. "This pursuit involves a bit of a leap of faith," says Drake Eggleston, director of computational and structural sciences at SmithKline Beecham's main UK research and development labs in Harlow. But he is convinced of the concept's long-term importance. "Structure-based drug design has developed more slowly than some had thought it might, but it is coming into its own and will be even more important in the future, aided by the availability of high-resolution macromolecular structures."

Pharmaceutical companies also air other concerns, for example that many of the publicly funded initiatives do

not involve human proteins or that many of the most interesting drug targets are not, at the moment, amenable to high-throughput approaches. The latter includes proteins that are embedded in cell membranes many of which are involved in signal transduction. Such proteins are the targets for half of the drugs currently on the market, but they do not dissolve easily in the aqueous solutions needed for NMR spectroscopy or for the production of protein crystals for X-ray diffraction studies (see main story).

Industry's hesitancy about how far to go in supporting the structural genomics consortium has bred similar reserve among some academics. "It is hard to come up with an idea of what we could offer the consortium until we know with more certainty its funding intentions," says Udo Heinemann, scientific spokesman for the Protein Structure Factory in Berlin.

But the chicken-and-egg situation will soon be resolved, claims Williamson. He is confident that the consortium will be up and running early next year, and will fund academic centres or small biotechnology companies for periods of at least three years.

time', above). But a series of task forces established at the First International Structural Genomics Meeting held in April in Hinxton, near Cambridge, UK, is now considering how overlap can be minimized and cooperation maximized. And once things take off, protein structures will roll in thick and fast. The target throughput for the German and Japanese initiatives is 100–200 proteins per year; for the NIH initiative, 10,000 new domain folds over ten years. "It will be at the very minimum five years before structural genomics starts to have a big impact on biology, and then it will explode," predicts Yokoyama.

For today's structural biologists, used to determining protein structures in their own labs, the idea of most of this work being sucked into the new protein structure factories may seem disconcerting. But the enthusiasts for structural genomics say that the really difficult work — deciphering biological functions from the outpouring of struc-

tural information — will still need to be done. And that will place insightful and creative structural biologists in hot demand. "Factory science does not spell the end for structural biology," stresses Heinemann. "Things will be even better."

Alison Abbott is Nature's Senior European Correspondent.

1. Perutz, M. F. *et al.* *Nature* **185**, 416–422 (1960).
2. *Nature Struct. Biol.* **7**, Suppl., 927–994 (2000).
3. von Itzstein, M. *et al.* *Nature* **363**, 418–423 (1993).
4. Kim, E. E. *et al.* *J. Am. Chem. Soc.* **117**, 1181–1182 (1995).

Web links

Protein Data Bank

♦ <http://www.rcsb.org/pdb>

Protein Structure Factory, Berlin

♦ http://userpage.chemie.fu-berlin.de/~psf/ifv_psf.htm

NIH Structural Genomics Initiative

♦ <http://www.nigms.nih.gov/funding/psi.html>

First International Conference on Structural Genomics

♦ <http://www.nigms.nih.gov/news/meetings/hinxton.html>

International Conference on Structural Genomics 2000

♦ <http://icsg2000.riken.go.jp>

Farming will only be sustainable when local people are truly involved

Sir— Yaalon in Correspondence (*Nature* **405**, 993; 2000) refers to my earlier Correspondence in which I argued that local farmers could feed Africa if they were given the chance (*Nature* **404**, 431; 2000).

I agree with Yaalon that available data on land resources are not always reliable. Their accuracy depends on data provided by national authorities. The UN Food and Agriculture Organization, which collects and publishes this material, is the only source for this worldwide information and is not to blame for any inaccuracies.

The comparatively low agricultural production in Africa today is not due to population pressure, the subsequent shortening of the fallow period and the depletion of soil nutrient status, as is generally believed. As I pointed out, there is a growing shortage of labour in rural areas as people migrate to the cities, and hence a reduction in cultivated acreage.

Technical solutions are well known: for many years agricultural development projects have focused on ways to increase production including improved soil management, intensification and diversification of the cropping pattern, adapted farming systems, restoration of the soil nutrient status, and so on. That these have not been implemented is often owing to socio-economic or cultural constraints.

Fertilizer is a good technical solution to the problem of poor soils, for example. But for local farmers, it still represents an extra cost which most cannot pay.

Yaalon's suggestion that "current subsidy and overseas aid should be used to provide the fertilizers and improved planting materials that are needed" is not a sustainable solution, as it replaces the structural problem of food aid with another structural problem of fertilizer aid. Once the free supply stops, farmers return to their traditional system.

Yaalon's vision that "many locally trained soil and extension specialists living in the region are needed to transform the economy from one of small rural farmers to one of market food production" has not been successful in Africa. Almost all efforts over the past 20 years have now been dismantled or vastly reduced, despite efforts to involve local expertise.

Substantial rethinking about aid is needed, to switch from a top-down to a bottom-up approach. The key questions are how to increase the sustainability of technical assistance and how to involve local people in follow-up activities. Examples of the former include supplying

farm implements and creating a small credit system; setting up a cooperative network for marketing agricultural products; attempting to end the practice of denying access to land by women, and so on. If the stakeholders are involved at all stages — ensuring that the project meets their demands — and their knowledge is included (updated if necessary), they are very much more likely to continue the activities after the project ends.

Yaalon's dedicated soil scientists might be helpful, not necessarily because of their technical skills but as facilitators for local people, to understand their problems and to use their advice. The soil scientists will learn that local farmers are usually well aware of their problems, and even know how to solve them, but generally lack the external support.

In my opinion, increasing food production in Africa is primarily a question of incentives and marketing assistance to farmers' communities. Once farmers are guaranteed a reasonable return for their work they will engage in efforts to produce more. With the extra earnings they will buy fertilizer, and break the vicious circle of subsistence farming.

Willy H. Verhey

Fund for Scientific Research Flanders, Department of Geography, University of Gent, Krijgslaan 281, B-9000 Gent, Belgium

Guidelines work better than animal welfare law

Sir— Your Opinion article, "In defence of animal research" (*Nature* **407**, 659; 2000), incorrectly states that the Federation of American Societies for Experimental Biology (FASEB) urges dispensing with a definition of distress for animals used in research, testing and teaching.

In a letter sent to the US Department of Agriculture (USDA) on 18 October, FASEB states that guidelines for the recognition of distress would be useful, but should not be included in USDA regulations or policy manuals because a single, standardized definition of distress would not help institutions to recognize, minimize or report animal distress across species and situations.

FASEB believes that local Institutional Animal Care and Use Committees (IACUCs) should take responsibility in this area and foster a partnership among scientists, veterinarians, veterinary technicians, animal husbandry staff, government and professional associations.

Rather than being a reactionary argument, as asserted in your editorial, the FASEB position is based on a careful review of data from the scientific study of pain

and distress in animals, and discussions with laboratory animal-care specialists from various disciplines including physiology, psychology and neuroscience.

Further, FASEB's recommendations are designed to protect the animals that scientists are indeed privileged to use. They acknowledge that a local IACUC is the best qualified to make decisions about the animals in its care, given their intimate knowledge of multiple factors affecting an animal's well-being.

Although FASEB cannot claim that all researchers are united in their opposition to including rats, mice and birds under the Animal Welfare Act, it represents 21 research societies with more than 60,000 members. Your editorial mentions Public Health Service regulations: more than 90% of the rats, mice and birds used for research in the United States are already covered by voluntary accreditation and/or Public Health Service policy. The research community is therefore objecting to redundant regulations, and not the principle of regulating the use of animals.

FASEB believes that modifications to USDA regulations should benefit animals, promote research and reduce administrative cost and regulatory burden. Including rats, mice and birds under the Animal Welfare Act would not accomplish these goals.

FASEB's complete statement is at <http://www.faseb.org/opar/news/docs/usda9x22.html>. We encourage dialogue with the scientific community on the evolving standards for determining pain and distress in laboratory animals.

Mary J. C. Hendrix

FASEB, 9650 Rockville Pike, Bethesda, Maryland 20814-3998, USA

Wrong signals about alliance's scope and aim

Sir— I would like to correct an error in your Opinion article "Systems Biology's Multiple Maths" (*Nature* **407**, 819; 2000) about the approach and the scope of the Alliance for Cellular Signaling. Our focus is limited to G protein-regulated and related (interacting) cellular signalling systems, not "every protein-protein interaction in two cell types". We seek quantitative information about these systems at all levels — not just detection of interactions between proteins. Further details can be found on our website, <http://afcs.swmed.edu>; see also *Nature* **402**, 219; 1999.

Alfred G. Gilman

Department of Pharmacology, University of Texas Southwestern Medical Center, Dallas, Texas 75390, USA

Helping western forests heal

The prognosis is poor for US forest ecosystems.

William Wallace Covington

The dry forest ecosystems of the American West, especially those once dominated by open ponderosa pine forests, are in widespread collapse. We are now witnessing sudden leaps in aberrant ecosystem behaviour long predicted by ecologists and conservation professionals (see *Nature* **407**, 5; 2000). Trends over the past half-century show that the frequency, intensity and size of wildfires will increase — by orders of magnitude — the loss of biological diversity, property and human lives for many generations to come.

Population crashes

Forest ecosystem health has declined pathologically ever since European settlement. Populations of native species have crashed while those of exotic species have irrupted. Indigenous cultural regimes and low-intensity disturbance regimes, such as surface fires, have been lost. Meanwhile, novel cultural practices, such as overgrazing and fire suppression, and unprecedented disturbance regimes (such as crown fires) have become increasingly dominant.

How can we help these ecosystems recover in a way consistent with their evolutionary environment, while enhancing opportunities for continued human use? Secretary of the Interior Bruce Babbitt and Agriculture Secretary Dan Glickman recommend extensive preventive fuel treatments such as thinning and prescribed burning. In their report to President Bill Clinton in September they requested substantial budget increases over the next 10–15 years to accomplish these. Congress responded by adopting many of the recommendations and increasing funding exponentially for fiscal year 2001; the president signed the legislation for that funding in October.

This new policy attention is critically needed. But focusing on abating crown fires without attention to the general ecological degradation of western forests is akin to treating a symptom and not the disease. The disastrous wildfire season of 2000 can be looked at in the framework my colleagues and I have developed for forest-fuel treatment. We have revised the current single-crop agriculture strategy to create an ecosystem-oriented approach that does not deviate too far from natural conditions (see Box, overleaf). The debate on forests must be driven not by ideology but by general agreement on the goal of coexistence with wildland ecosystems. The starting point must be good science. To approach the problem any other way will open the door for management decisions that



Old growth ponderosa pine in a restoration site (top) and after ecological restoration treatment.

may lead to short-sighted, uninformed 'tree-whacking' and forest burning that will only exacerbate the problem.

I am not so naïve as to believe that science will drive the entire debate. But, having spent 25 years trying to advance the need for ecosystem restoration, I know that debate descends to paralysis when it begins with ideology. In the little time left, it is imperative that fuels treatments should be driven by ecological restoration principles; that scientists should actively inform the debate; that scientists, managers and the public should not be afraid to use what we currently know to identify restoration treatments; that all restoration should be conducted in an explicitly formal, adaptive management context; and that multiple approaches should be tested and carefully monitored.

Restoration principles

Ecological restoration offers a practical approach for developing scientifically and ethically sound fuel-reduction treatments which not only treat wildfire symptoms, but also attack the underlying causes of ecosystem health decline. Restoration is mandated for degraded areas set aside as natural areas or wilderness, but it is also a desirable goal in the management of lands where ecosystem health

and resource use are shared goals. Finally, scientifically rigorous and adaptively designed restoration plans can offer a common ground to resolve preservation/use conflicts.

What is needed urgently is solid scientific evidence. Researchers, conservation professionals and other knowledgeable people concerned about western forest ecosystem health must agree and publicize an objective standard for information to counter the tendency for propaganda and ideological 'spinning' of research results used in developing treatments or to drive policy discussions.

Some of the stakeholders in forest management — the forestry industry, environmental groups and other resource users — use misrepresentation and pseudo-science to justify their positions and undermine the credibility of others. I have seen our restoration treatments at Flagstaff intentionally misrepresented on the one hand by some environmental activists as monolithic in approach and scientifically unsound, and on the other by some representatives of the wood and grazing industry as a justification for returning to old-time commodity resource production from federal lands. The reality is that ecosystem health comes first: we test multiple hypotheses operationally and vary treatments depending on site conditions. We see resource utilization as a way of accomplishing restoration, not as a goal.

Because I believe strongly that restoration must begin immediately I have been criticized by colleagues who argue that more information is needed first. I strongly disagree with this because the current crisis in

western forests demands action now; there is sufficient information to design and implement scientifically valid approaches that will help determine how best to proceed.

Sound knowledge about restoration is available, based on the abundant scientific research that began in the 1890s and continues today. We have solid information about forest conditions before European settlement, changes in fire regimes, deterioration of over-all ecosystem health, and ecological responses to thinning and prescribed burning—the key elements of any attempt to restore ecosystem health in ponderosa pine and related ecosystems. Both early naturalists' data and contemporary research show that today's overcrowded stands of trees do not sustain the diversity of wildlife and plants that existed a century ago. Open, park-like stands that have not been severely disrupted by fire suppression have greater plant, insect and bird diversity than stands that have become overstocked during decades of fire exclusion. These open, park-like stands also show greater old-growth tree vigour and resistance to insect attack.

Stopping ecologically based forest restoration because of an ideological opposition to tree cutting is not saving forest ecosystems, as some would like to believe, but only contributing to their demise and causing severe losses to the wealth of species that depend on them. It also places ecosystem health practitioners in a situation similar to eliminating surgery from the options for treating human diseases.

Various ecosystem restoration options are being investigated at research sites across the American West. Most apply and test treatments developed locally by scientists, managers, environmental activists, resource users and members of the public. But it is also imperative that we accept the responsibility to apply the extensive knowledge we already have, before more forests are lost. Ecosystems have been highly fragmented and degraded by decades of overuse. Restoration is not necessarily simple, nor is success always guaranteed. But restoration-management approaches are superior to the destructive effects of unnaturally intense fires. The risks of inaction far outweigh the risks of such treatments.

I do not advocate a 'one size fits all' approach to restoring the ecological integrity of ponderosa pine forests. Management approaches should suit the place that needs restoring, its pre-European settlement reference condition and its relationship to the broader ecosystem and the communities that live within it. In this sense, ecological restoration is neither a strict recipe nor a rigid set of prescriptions, but is an adaptive process for restoring and enhancing ecosystem health and sustainable human uses of the land.

We must change our approach to fuel and forest management. The shocking fire season of 2000 and the environmental and economic damage left in its wake are tragic consequences of management and interest groups that have



Under control: a prescribed burn at Flagstaff.

failed to understand natural ecosystem structure and function. Virtually all our scientific knowledge is based on studies of degraded ecosystems—those that have been extensively altered by fire exclusion, industrial exploitation and other exotic cultural practices.

As the new US government takes shape, now is the time for land-management agencies, policy-makers and stakeholders to identify what is needed to achieve ecologically healthy forest ecosystems. The discussion should start with solid science, embrace a long-term ecological perspective and then incorporate social and ideological issues. Scientists must make this happen by actively promoting what we know and the ways this knowledge should be applied. The current crisis requires that we act immediately, using an adaptive environmental assessment approach.

Today's wildfires are so extreme in their behaviour and effects that they are in many ways worse than clearcutting. Critical habitat for threatened and endangered species is destroyed, watershed function is disrupted and human habitat value reduced for centuries to come. And such wildfires are a threat to human lives and property. To act now is to save the patient. To act now means a healthy, biologically diverse forest that is an asset, not a threat, to future generations. Is there really an alternative? ■

William Wallace Covington is at the Ecological Restoration Institute, Northern Arizona University, PO Box 15018, Flagstaff, Arizona 86011, USA.

► <http://www.for.nau.edu/ecorest>

Ecological restoration

Ten years ago, Carl Walters and Crawford Holling memorably recommended "large-scale management experiments and learning by doing" (Walters, C. J. & Holling, C. S. *Ecology* **71**, 2060–2068; 1990). We have pursued this approach in the 'Flagstaff Plan'. In collaboration with partners in the environmental community, conservation practitioners and interested parties from government and a range of organizations, my colleagues and I at the Ecological Restoration Institute at Northern Arizona University have developed a general framework for ecologically based restoration treatments.

Scientific framework

Ecological restoration is the restoration of natural ecosystem structures and processes. Treatments are based on reference conditions (the evolutionary environment context) and cover evolutionary and conservation biology, and ecosystem ecology principles.

Social and political framework

In an ecosystem ecology approach, social and political concerns play a major part in defining treatments. So stakeholders must be engaged, especially community-based partnerships linked to regional and national agencies and interest groups, with policy-makers, natural-resource specialists and resource managers.

Operational framework

Financial and personnel constraints place geographical limits on treatments. So emphasis is placed on strategically located restoration fuel breaks that are anchor points for large, landscape-scale treatments. These breaks can protect key landscape ecosystem components such as human communities, critical habitat for threatened or endangered species, and core areas of greater ecosystems such as wilderness areas and national parks.

Ecosystem-management framework

Restoration and fuel-reduction goals should be integrated with overall ecosystem conservation and management goals. Reference conditions serve as a starting point for the goal of scientifically based land-management objectives.

Economic framework

Economic analyses should consider all costs and savings. Restoration-based fuel treatments save money by avoiding the costs of firefighting, rehabilitation and compensation for property damage. They are also an investment in protecting lives. They present new opportunities for rural economic development through restoration-related jobs and products. Ecological economic analysis will probably indicate that benefits greatly outweigh costs.

Ethical framework

We have a responsibility to future generations to solve ecosystem health problems. Ecological restoration speaks to the land ethic—humans should be good stewards and show a caring concern for nature.

Anthropologists under fire

Charges of abuse of a remote tribal people must be answered.

ROBERT HARDING

Darkness in El Dorado: How Scientists and Journalists Devastated the Amazon

by Patrick Tierney
W. W. Norton: 2000. 416 pp. \$27.95

Robert N. Proctor

Scandal is nothing new to anthropology — think of the furore surrounding the discovery that Bronislaw Malinowski referred to the Trobriand Islanders as “niggers” in his field notes. But Patrick Tierney’s book makes previous such disputes seem pale by comparison. The book has already ignited a firestorm of controversy among anthropologists, and although it will take some time to evaluate all of the evidence, the charges levelled in this book — which include sexual abuse, falsification of data, collaboration with illegal mining and criminal medical neglect — rank as the most serious with which the anthropological profession has been charged.

The Yanomami people of Venezuela and the Brazilian Amazon were a relatively isolated group when they first captured the attention of anthropologists and atomic scientists in the 1950s. In 1958 Marcel Roche of Venezuela’s Instituto Venezolano de Investigaciones Científicas was paid to carry out a series of experiments involving the administration of radioactive iodine-131. His task was to investigate why highland populations such as the Maquiritare suffered more from goitre than did lowland Indian populations. The US Atomic Energy Commission (AEC) funded these and other experiments, including several involving the injection of radioactive iron, as part of a broader campaign to study the medical effects of nuclear war. Tierney links these efforts to the larger AEC programme of “body-snatching” — sampling bone and other body parts to determine the rates at which radionuclides become fixed in the human body.

The AEC was also one of the major sponsors of studies on the Yanomami by anthropologist James Neel and his graduate student Napoleon Chagnon. One of Chagnon’s jobs was “to prepare thousands of Yanomami at dozens of villages to receive teams of scientists”. Their goal was to measure mutation rates in this supposedly pristine population, unexposed to modern civilization. The AEC funded Neel’s Yanomami work from 1965 to 1972 as part of its larger, multi-million-dollar mission to determine “the mechanisms by which radiation induces changes in the genetic material of cells”. Neel and his colleagues took thousands of blood samples from the Yanomami, in order to identify



Distorted lives? Were ‘women’s activities’ ignored in researchers’ desire to record violence in the tribe?

mutations that might clarify “the health risks associated with exposures of people to energy-related radiation and chemicals”.

The Yanomami came to the attention of a much wider audience in the late 1960s, when Chagnon published his book *Yanomami: The Fierce People* (Holt, Rinehart & Winston). It quickly became the most popular anthropology text ever, with several million copies printed. With funding from the AEC, Chagnon orchestrated and directed films such as *The Ax Fight* and *The Feast*, portraying the Yanomami as inherently violent and proof that (as Tierney puts it) “ruthless competition and sexual selection cannot be legislated away by idealistic do-gooders”. Chagnon himself was explicit about his desire to wipe away “all the crap about the Noble Savage”.

A central thesis of Tierney’s book is that much of the violence found among the Yanomami was contact-induced. Anthropologically induced, in fact, because of a ‘metal rush’ that was inadvertently launched for the steel goods — especially fish-hooks and machetes — that Chagnon and his colleagues were doling out in return for various favours.

Tierney claims that anthropologists altered the economic and political balance of Yanomami society with their elaborate gifts (“checkbook anthropology”) and the ways they gathered information. Chagnon would often ask for the names of Yanomami ancestral dead, for example, to determine genetic relatedness. This violated one of the

strongest Yanomami taboos. Locals were often unwilling to cooperate, but members of rival villages could sometimes be cajoled into supplying the names, which tended to exacerbate inter-village tensions. Tierney says that the word ‘anthro’ entered the Yanomami vocabulary with a meaning “something like the opposite of its original Greek meaning” (*anthropos*, human), and that the Yanomami came to think of an ‘anthro’ as “a powerful nonhuman with deeply disturbed tendencies and wild eccentricities — an Olympian in a funk”.

There are also charges of sexual abuse. Tierney details the sexual frolicking of Jacques Lizot, a French anthropologist, who, in one Yanomami village, ruled over a harem of young males with whom he routinely traded steel goods for sexual favours — two sex acts for a machete, six for a shotgun, and so forth. Lizot’s dalliances entered the Yanomami lexicon, so the Yanomami word for anal intercourse became *Lizo-mou*: ‘to do like Lizot.’ Lizot in his writings portrayed the Yanomami as “sexual innovators of stunning sophistication, an Erotic people” — a distortion reminiscent of that which Derek Freeman exposed in the work of anthropologist Margaret Mead, whom he accused of idealizing the sexual freedoms of her Samoan informants.

A further charge is that some of the fights filmed for the famous *Ax Fight* film and various other ethnographic films were staged by



Meddling missions? It is claimed that the economic and political balance of tribal society was altered.

Neel, Chagnon and film director Timothy Asch to buttress their views about violence in Stone Age society. Asch himself admitted in 1992 that certain aspects of the fights were embellished — in the dramatic sequence involving a head wound in *The Ax Fight*, for example, the grisly 'thud' accompanying the injury was provided by striking a watermelon.

Tierney describes a formula for Yanomami film-making: "the way to make a successful Yanomami movie was to build a new shabono [village round-house], sponsor a feast, create a new military alliance, and record a raid by the newly created power." He adds that a frequent (and never filmed) sequel was an epidemic — caused by the spread of contact diseases such as measles, to which these isolated people had very little resistance — that might kill a quarter or more of the Yanomami actors.

Tierney follows Brian Ferguson, author of *Yanomami Warfare: A Political History* (School of American Research Press, 1995),

in arguing that important aspects of Yanomamo health and well-being were often neglected or worsened by the film-makers. None of the 20-odd documentaries made about the Yanomamo community of Mishimishimabowei-teri, for example, mention the fact that the village was decimated by measles and malaria shortly after the filming in the early 1970s.

Women's activities were, by and large, ignored: Chagnon reportedly would not let Asch train his camera on anything but aggressive behaviour. So, when Asch urged him to record some everyday women's activities, Chagnon "whipped around" and asked: "What makes you think there are any women's activities?"

Many of the people filmed were in desperate need of medical attention, but Tierney suggests that the 'no-treatment' policy begun at the Atomic Bomb Casualty Commission in Hiroshima — which Neel headed — was carried over into the Amazon, with deadly consequences.

During the filming of *Warriors of the Amazon* by the television company Nova, for example, a camera broke and a substitute was flown in from London. But a dying Yanomami woman featured in the film couldn't be ferried an hour downriver to save her life. The woman's slow death over a period of a week was diligently filmed, a process Tierney describes as involving "a deep, unconscious conflict of interest between the film's need for a climax [the woman's death, along with that of her child] and the woman's need for emergency evacuation". Tierney argues that although neither Chagnon nor Neel was involved in the making of this particular film, the emphasis on violence followed a narrative model established by the earlier films.

One of the most controversial and shaky charges in the book is that the deadly measles outbreak of 1968 may have been started by Neel as part of a grotesque experiment to test the resistance of primitive populations to contact with infections. The circumstances surrounding the outbreak are unclear, and

Tierney's claim that the vaccine chosen for the inoculations — the Edmonston B live vaccine — was "contraindicated" seems to rest on rather flimsy evidence. It is true that the Venezuelan government was already using the allegedly safer Schwarz vaccine in its national campaign against the disease, but Tierney presents no convincing reason to believe that Neel thought it might set off an epidemic — if that is, in fact, what happened. (Neel and his entourage may have inadvertently helped spread the disease, although the author's account of a possible mechanism seems rather muddled.)

In the flurry of e-mail exchanges responding to Tierney's book before its publication, Susan Lindee, a science historian at the University of Pennsylvania, showed that Neel consulted the US Centers for Disease Control and Prevention in Atlanta before deciding to use the Edmonston B vaccine. And judging from Neel's field notebooks, which are preserved at the American Philosophical Society, he seems to have been genuinely disturbed when the measles outbreak occurred.

It will take some time to sift through all of Tierney's claims. He may well have exaggerated some of the harm inflicted on the Yanomami by atomic anthropologists. But even if only a fraction of his charges can be verified, the book could be of significance in reopening debates about the impact of atomic-energy priorities on scientific practice. President Bill Clinton's Advisory Committee on Human Radiation Experiments focused almost exclusively on experimental abuses on American soil; little was done to explore misdeeds overseas. Reading Tierney's account, one is prompted to ask what other abuses might exist that we don't yet know about — in the former Soviet Union, for example, or in some of Africa's anthropological stamping-grounds.

Equally important may be the implications for ongoing disputes over the nature and limits of 'informed consent'. Do scientists have a right to use blood samples collected at a time when donors were not

New in paperback

Rosalind Franklin and DNA

by Anne Sayre

W. W. Norton, \$13.95

Tides: A Scientific History

by David Edgar Cartwright

Cambridge University Press, £18.95, \$29.95

"Now comes David Cartwright's *Tides: A Scientific History*, a major contribution to the history of astronomical and geophysical sciences, filling that serious gap in the scientific literature ... This book is a masterpiece." Paul Melchior, *Nature* **398**, 119–120 (1999)

A Cursing Brain? The Histories of Tourette Syndrome

by Howard I. Kushner

Harvard University Press, \$16.95

"This book is a 'must' for anyone interested in the history of medicine, neurology and psychiatry as well as Tourette syndrome. There is no doubt that this is the best exposition of the syndrome's history in the literature ... Kushner's is one of the most exciting and intriguing textbooks that I have read: clinically and historically correct, extremely well written and an erudite and scholarly treatise." Mary Robertson, *Nature* **400**, 422–423 (1999)

Almost Like A Whale: The Origin of Species, Updated

by Steve Jones

Anchor, £8.99

"This 'Origin of Species — Revised Edition' may lack the sense of danger and discovery of its predecessor, and the radicalism has been replaced by a presumption of the truth of natural selection. Yet Jones succeeds impressively in his desire to bring us up to date on the facts of evolution: to read this book is to see in one place much of the sweep and grandeur of evolution by natural selection." Mark Pagel, *Nature* **401**, 853–854 (1999)

informed of their eventual use? How should we deal with biological material gathered under suspect circumstances? Does it matter whether or not the Yanomami of today and of 30 years ago would approve of their blood being used to test whether 'headmen' leave more offspring?

However we answer such questions, it is important that they be raised. Tierney has opened up a rather large can of worms and, judging from the attention the book is getting, there is a lot of explaining to do. Anthropologists in the 1960s and 1970s turned Yanomamo territory into a vast proving ground for ideas about sociobiology, radiation injuries and the naturalness of human aggression, and the fallout from that experience has yet to settle. ■

Robert N. Proctor is in the Department of History, Pennsylvania State University, University Park, Pennsylvania 16802, USA.

Across the industrial divide

The Great Divergence: China, Europe and the Making of the Modern World Economy

by Kenneth Pomeranz

Princeton University Press: 2000. 382 pp. \$39.95, £25.95

Anthony Vice

In his seminal work on Britain's Industrial Revolution, *The Workshop of the World* (1961), J. D. Chambers described how an American visitor, touring the textile towns and villages, exclaimed over the workmen's cottages, with their stone fresh from the masons' yards. Kenneth Pomeranz, an up-to-date American visitor, reviews this period ('industrial revolution' is not a term he appears to favour, preferring 'sustained industrial growth') from the perspective of the Far East.

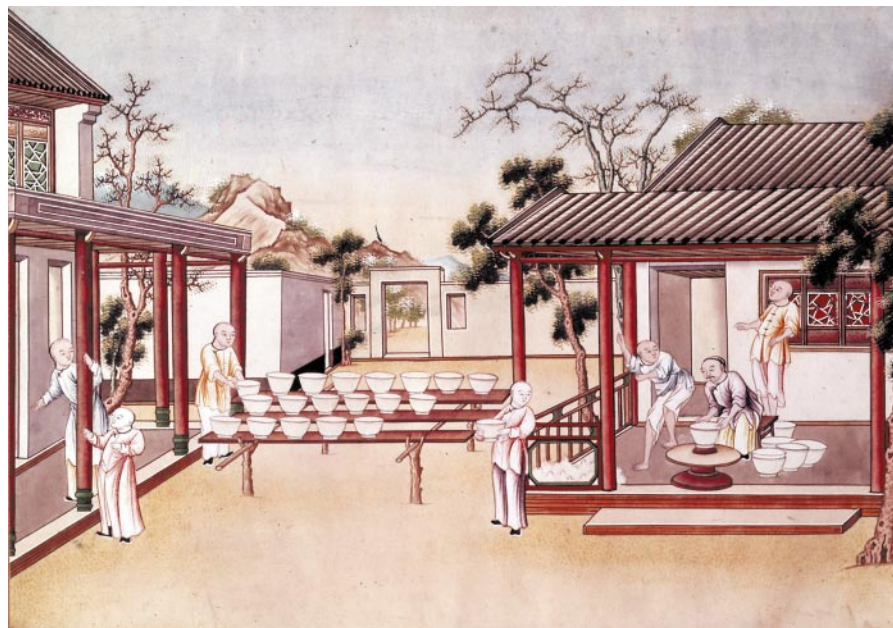
Briefly, his thesis is that the conditions for an industrial revolution existed in the Yangtze Delta of China, in Japan and in Gujarat in India. Population levels matched those of Europe, whereas Asian techniques in agricultural land management and the use of fuel, Pomeranz suggests, were superior to those in Europe. But as we all know, Europe went on to enjoy an Industrial Revolution; the Asian economies hit a cul-de-sac. The East Asia hinterlands boomed after 1750, but this local growth inhibited those regions from exporting to the faster-growing Yangtze Delta. As a result, growth at the core of the East Asia economy essentially stopped, and what growth did take place was forced along labour-intensive paths. That represents the 'great divergence'.

My immediate reaction is to applaud the novelty of this analysis. Only 10 years ago it would have been difficult to find literature to support Pomeranz's thesis in relation to China; 25 years ago it would have been hard

to make such a case in respect of Japan.

The self-evident question is why Europe and Asia performed so differently — why was there this divergence? Pomeranz's answer is essentially twofold: first, Europe enjoyed large coal resources, which were accessible and therefore cheap. And second, it benefited from what Pomeranz calls "the fruits of overseas coercion", by which he means slavery. "The slave trade ... enabled Europe to exchange an ever-growing volume of manufactured exports for an ever-growing volume of land-intensive products."

These arguments seem to offer both too little and too much. On the importance of slavery, Pomeranz quotes an estimate that the fruits of overseas coercion could not have been responsible for more than 7% of gross investment by late-eighteenth-century Britons, and for Europe as a whole the figure would have been far less. Coal of itself was of limited relevance to the Industrial



Porcelain-making: sophisticated production, but not coal, iron and steel, which dominated in Europe.

New in paperback

Sudden Origins: Fossils, Genes, and the Emergence of Species

by Jeffrey H. Schwartz

Wiley, \$18.95, £14.99

"Schwartz presents a detailed and informative historical account of evolutionary biology. In fact, the book could be read as a history of evolution and will probably occupy such a niche. Schwartz is much more ambitious than this, however: he wants to convince the reader that his "new evolution" is the Holy Grail of the field." Eörs Szathmáry, *Nature* 399, 745 (1999)

The Fabric of the Heavens: The Development of Astronomy and Dynamics

by Stephen Toulmin & June Goodfield

University of Chicago Press, \$15, £10.50

Edison: A Life of Invention

by Paul Israel

Wiley, \$18.95, £12.50

"Paul Israel, in this latest biography, has done a remarkable job. Not only has he given us fresh insights into a complex personality, but he has set this against the backdrop of a dramatically changing American society driven on

remorselessly by the second Industrial Revolution in which Edison was a pivotal player." Willem Hackmann, *Nature* 397, 34–35 (1999)

Pierre-Simon Laplace (1749–1827): A Life in Exact Science

by Charles Coulston Gillispie

Princeton University Press, \$19.95, £12.50

"Gillispie's distinguished biography is a magisterial survey of one of the most influential scientists of the past two centuries." William R. Shea, *Nature* 391, 855–856 (1998)

Revolution, being mainly used for domestic fuel. What mattered was the alliance of coal, iron and steel, which was given sharp impetus by the Seven Years' War and later again by the campaigns against Napoleon. British iron- and steel-making skills, based on earlier manufacture of clocks and guns, were of a high order. One indicative story relates how the German industrialist Alfred Krupp took an assumed name and lived in Sheffield to learn the steelmakers' skills.

There is something of a paradox in the thesis of this cogent and well-researched book. A great divergence between the economic performances of Asia and Europe would be remarkable only if there had been a widespread expectation of convergence. There is little evidence that the Industrial Revolution had been expected by economists of the time to reproduce itself in the East. That is not because China, Japan and India were viewed as remote countries of which we know little, but rather because there were fundamental differences in the structure of the two regions.

First, and probably most important, were the very different legal systems. The German economist Max Weber argued that a rational legal system was necessary for the smooth functioning of a capitalistic economy. A recent study by Hernando de Soto, seeking to explain why capitalism has proved so effective in North America and Western Europe but failed to deliver in Russia, South America and Africa, points to the absence of legally enforceable property rights. Second, the Industrial Revolution was a British rather than, as Pomeranz tends to argue, a European phenomenon. He himself points out that European industrialization was still quite limited outside Britain until at least 1860 — remarkably late by UK standards.

Until the mid-nineteenth century, a small island to the north-west of continental Europe (Scots and Welsh contributed, as well as English) produced a crop of entrepreneurs possessing financial skills allied to outstanding qualities in mechanical engineering. Many were harsh, demanding and narrow-minded, as we know from the literature; but they were also honest — although one George Hudson, the subsequently disgraced 'railway king', was the exception.

Their successors fared less well, later in the nineteenth century, in the age of electrical power and the internal-combustion engine. That was another great divergence — between Britain and continental Europe, in Britain's favour, until around 1860, and from then onwards between Britain and North America and continental Europe, to Britain's disadvantage. It's a moot point whether or not this latter divergence continues.

Anthony Vice can be contacted at the United Oxford & Cambridge University Club, 71 Pall Mall, London SW1Y 5HD, UK.

Science in culture

The body spectacular

Skeletons in the medico-artistic cupboard

Roy Porter

One of the things you first notice about the people attending *Spectacular Bodies*, a quite amazing exhibition at London's Hayward Gallery, is the complete mix it has drawn in, rather than just your usual arty crowd. And that's not surprising, for we're all fascinated by our physical reality, our flesh and blood. But what a paradox is embodied there!

Our bodies, after all, are what we're most familiar with, yet they are also the heart of darkness, all unknown. Except occasionally, by courtesy of medical technology, we never see our insides. Save when we're in pain, we never feel most of our organs. And, religious or not, we sense there's something sacrosanct, taboo — or at least profoundly private — about our innards: most of us are squeamish enough about having our outsides seen naked. And any reminder of mortality is deeply disturbing.

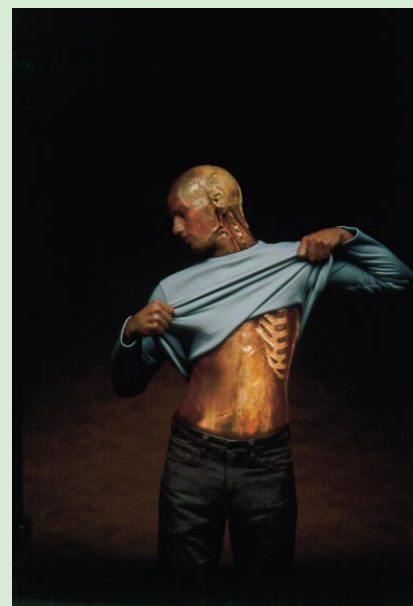
Hence the shocking nature of the traditions of scientific art and artistic science represented in the exhibition. From the Renaissance on, an unholy alliance of anatomists and artists began to dissect the human body and turn what they saw into artworks: drawings, paintings, écorchés (flayed muscle-men torsos, often holding flaps of skin aside), huge sculptures, little figurines and wax models that could be taken apart, layer by layer.

And there is something creepy besides. When a Dutch anatomist pickled a human fetus, what made him think to attach beads to its tiny hands and wrists? When wax modellers carved a torso, why did they affix curly tresses to the head, or add pubic hair, or give the faces individualized expressions? Something prurient, even pornographic, seems to be at work.

Doubtless, the draughtsmen and craftsmen had their reasons: artists did dissections so as to create objects of beauty; anatomists wielded the scalpel the better to know their pathology and be able to heal. But the question of voyeurism inevitably arises. What do we make of the pleasure that artists and anatomists so evidently derived from examining and dismembering the dead? And what of our own enjoyment?

We see, in the frontispiece to his *De Fabrica* (1543), the young Vesalius proudly displaying the inner secrets of the woman he has just dissected, most likely a criminal. Two centuries later, Hogarth presents us with a parody of that scene in his *Four Stages of Cruelty*, hinting that the true criminals are the surgeons — or perhaps us, the viewers.

Likewise, a vast canvas by André Brouillet shows Jean-Martin Charcot, the 'Napoleon of the neuroses', displaying a 'hysterical' woman, lace blouse slipping down her breasts, in front of an assembly of students. Is this medicine or is it a sex show? This exhibition's power lies in forcing us, time and again, to confront our own reactions to the skeletons in the medico-artistic cupboard. All



An advertising image based on a wax figure by Clemente Susini, 1804.

credit to the curators, Martin Kemp and Marina Wallace, for their superb achievement.

Two aspects of the exhibition were mildly disappointing, however. The top floor is devoted to the scientific aesthetics of the face. But with a few exceptions — notably Franz Xaver Messerschmidt's grinning and grimacing physiognomical busts — that section carries less force than those featuring the body at large.

And the attempt to weave modern, specially commissioned items into an exhibition that is mainly historical seems a missed opportunity. There are too few contemporary pieces to add up to more than a token appearance. And with the exception of John Isaacs' life-sized auto-torso, the modern pieces seem both too knowing and rather conventional. Here lies a paradox indeed: modern art courts the reputation of being shocking — think of Tracy Emin and Damien Hirst — but it is actually the academic artists, from Leonardo and Rembrandt to Johann Zoffany and George Stubbs, who provide the greatest jolts.

On entering the exhibition, what you see first is a series of Dutch group portraits featuring dapper doctors gathered around semi-dissected corpses. Instruments in hand, some look as if they are about to carve Sunday lunch; others seem about to give the cadaver a manicure. There is something shocking in the detachment of both the artistic and the anatomical gaze. It is an unnerving opening to an unnerving exhibition, which is not to be missed.

Roy Porter is at the Wellcome Institute for the History of Medicine, 183 Euston Road, London NW1 2BE, UK.

Spectacular Bodies: The Art and Science of the Human Body from Leonardo to Now is at the Hayward Gallery, London, until 14 January 2001.

The sounds of silence

No news about cancer prevention is good news.

Robert C. Young

For centuries, hepatitis has exacted a devastating worldwide toll, not only from the ravages of the disease itself, but in the later development of cirrhosis of the liver and the nearly always fatal primary liver cancer. Now, as the new millennium begins, that devastation is abating. Thirty years ago, Baruch S. Blumberg discovered the hepatitis B virus and invented a vaccine that has protected millions. The vaccine has also brought about a dramatic decline in hepatocellular carcinoma in places where the disease was endemic, such as Taiwan. This withering away of what was once one of the commonest terminal cancers is happening quietly, without the publicity associated with the complex interventions we have come to equate with medical progress.

Cancer researchers have long sought effective prevention and screening methods, but, with the exception of tobacco avoidance, cervical smears, mammography and a handful of others, there have been few advances. Nevertheless, even these efforts have produced substantial benefits. For the first 90 years of the twentieth century, a progressively greater proportion of people died from cancer in the United States, as in most other developed countries. But US cancer mortality has been falling since 1990, as a result of advances in prevention, screening and treatment.

With the revolution in molecular biology and the development of tools to define genetic and environmental risks, to determine molecular targets for drugs and vaccines, and to identify at-risk populations, science has an unprecedented opportunity to make this new century the age of cancer prevention.

Although there have been dramatic advances in our understanding of cancer, and marked improvements in our ability to treat it, the disease continues to take a gigantic toll. Globally, more than 10 million people develop cancer each year and over 7 million will die because of it — almost 20,000 people each day. Although surgery, radiation and drug treatments will continue to be important, research is identifying a growing number of effective cancer prevention strategies.

Two years ago, tamoxifen, a drug com-



Unhealthy lifestyle: *Skull With A Burning Cigarette*, by Vincent Van Gogh.

monly used to treat breast cancer, was shown to reduce the incidence of that cancer in high-risk women by 49%. Within a year, scientists were trying to work out whether raloxifene, an osteoporosis drug, would produce the same benefit with fewer side effects — two more examples of discoveries for one purpose finding applications elsewhere.

More and more preventive agents are being evaluated. Aspirin, non-steroidal anti-inflammatory drugs, folate and oestrogens all guard against colon cancer. Celecoxib, a cyclooxygenase-2 inhibitor, appears to decrease the incidence of recurrent polyps in patients with familial adenomatous polyposis. These discoveries push us even closer to the identification of agents that will prevent or delay cancer development.

Evidence that particular subtypes of papilloma virus lead to premalignant lesions and cervical cancer has led to clinical trials using recombinant virus capsid proteins to induce immunity. Vaccinating high-risk populations of young women should reduce

the 250,000 deaths worldwide each year, 80% of which occur in developing countries.

Increasingly, cancer research centres are integrating many disciplines to focus on prevention — genetics, epidemiology, chemoprevention, drug and vaccine development, risk assessment, behavioural research and bioinformatics — to support a growing network of prevention-related clinical trials.

This shift towards prevention should, paradoxically, produce less fanfare than the medical breakthrough *du jour*. Just as in the past millennium each generation had difficulty perceiving the cost to their forebears of diseases such as smallpox, polio and syphilis, so the spectre of cancer should recede from public view. Success is to be measured by the absence of disease, and less and less by its cure. It is rarely news when nothing happens: to borrow a phrase from a Simon and Garfunkel song, "The words of the prophets are whispered in the sounds of silence". ■

Robert C. Young is at the Fox Chase Cancer Center, Philadelphia, Pennsylvania 19111, USA.

Ars longa, vita brevis

Last remaining astronomy department disbanded.

James Alan Gardner

9 November 2270: In a move described as “long overdue”, the Astronomy department of the University of California, Berkeley, has been dissolved. Faculty and students from the defunct department will be shifted into other areas of the university, particularly sociology and fine arts. Berkeley’s astronomy programme was the last of its kind at any major educational institution.

“We tried to keep the place going,” said former department chair Dr Jeremy Washburn, “but with Astronomy departments folding in so many other universities, we knew it was just a matter of time. The enthusiasm just isn’t there anymore.

“I confess,” Dr Washburn continued, “I understand how other people feel. Celestial mechanics really lost its charm when we learned all the interesting stuff is artificial.”

Dr Washburn was referring to revelations from advanced extraterrestrial races that they are personally responsible for prominent astronomical phenomena. For example, the Vingex of Betelgeuse claim to have created all the binary and trinary solar systems in our galaxy by dragging stars closer to each other.

“It makes for more attractive visual composition,” said Speaker 183-478D, cultural attaché to the Vingex embassy on Earth. “Which is more interesting: a single sun just sitting in the middle of nowhere, or a group of colour-coordinated suns that set each other off nicely against the black background?”

“And it’s even better when you add a few planets to weave in and out between the stars. The orbits aren’t stable gravitationally, but you can keep them in line with a...”. Here the Speaker’s translator implant whistled several times, representing an untranslatable word — presumably the name of a Vingex device for adjusting the orbits of wayward planets.

Later in the same interview, Speaker 183-478D repeated the Vingex’s offer to procure a binary companion for Sol, Earth’s own sun. While the Speaker declined to say which star might be chosen as a suitable partner for Sol — “we want it to be a surprise” — hints were given that the star is blue-white, well-established in the main sequence, and “seldom given to solar prominences or other unattractive outbursts”.

When asked to comment on the Vingex offer, Dr Washburn said it was just the sort of

thing that had taken the fun out of astronomy. “There was a time when we’d have been fascinated by a blue-white sun locked close to a yellow one like Sol. It would have been worth a few review letters and perhaps a PhD thesis. Now it tells us more about the Vingex’s aesthetic sense than it does about stellar evolution.”

The Vingex are not the only race involved in large-scale cosmological manipulation. It is well known, of course, that *Homo sapiens* first made contact with alien life-forms when a group of Pleonines arrived in our solar system to “freshen up” the rings around Saturn and other outer planets. According to the Pleonines, several millennia had passed since the rings last received any touch-up work; colours were badly faded, and visual appeal had suffered.

During the restoration work, the Pleonines demonstrated techniques for creating new rings, as well as simulating colours through diffraction and producing complicated “braid” patterns. They also added two more giant red spots to Jupiter’s atmosphere and moved Pluto four billion kilometres closer to Earth because humans were having trouble seeing it where it was.

“We enjoy giving others the chance to view our work,” the Pleonine Queen explained. Although members of her race are primarily interested in expressing themselves through planetary rings and giant atmospheric anomalies, Her Majesty has adorned the entire surface of Pluto with a portrait of herself, created by meteorite collisions. The portrait appears to be rendered in the style of the early Cubists; in

the Queen’s case, however, this may actually be photo-realism.

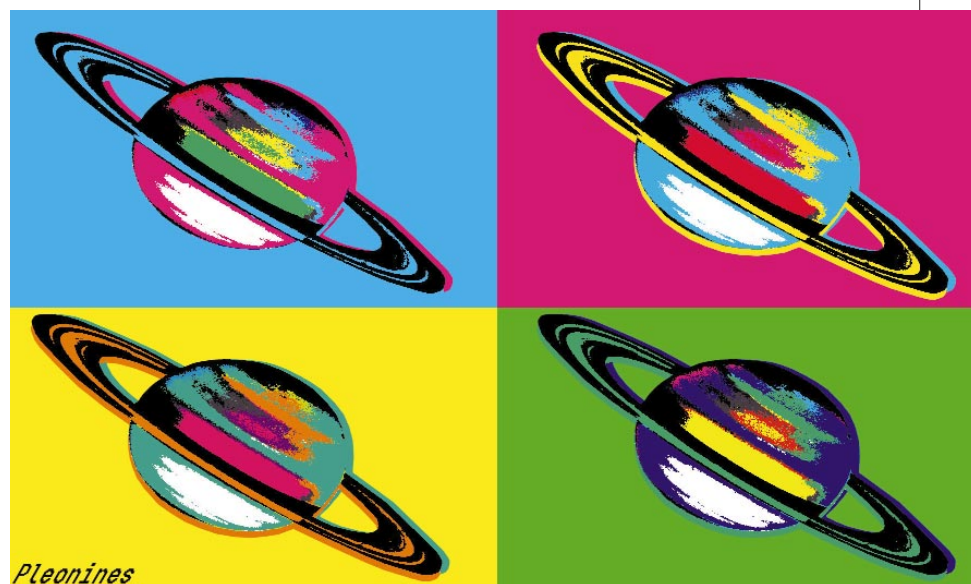
Once the Pleonines had ‘broken the ice’ by making contact with humankind, other aliens soon arrived to ask our opinion of their art. Notable amongst these visitors were the all-mechanical Regimoids, creators of every pulsar in the universe (“pulsars ... are ... regular ... pulsars ... are ... beautiful ...”), the Über-Masons who constructed the Great Wall, and the so-called Bangers responsible for supernovae.

“Oh yes,” said Dr Washburn of Berkeley. “I knew our department was in trouble when I heard about supernovae. We had all those great theories about stellar collapse ... then suddenly, we found out novae were just the work of ET punks who liked blowing things up. The very next day one of our best Astronomy professors transferred into Humanities. She still gives the same slide show she did in Astro 101, but now it’s called Art History.

“That was discouraging enough,” continued Dr Washburn, “but the thing that broke my heart was those jelly people showing up to take credit for the Horsehead Nebula. I can still remember their words: ‘My God, from this angle it looks *fabulous!*’” Dr Washburn sighed. “I used to think it did too.”

Shaking his head sorrowfully, Dr Washburn cleared his desk, left his office, and locked the door behind him.

James Alan Gardner studied applied mathematics at the University of Waterloo, and then immediately began writing science fiction instead. He lives in Kitchener, Ontario, Canada; his novels include *Commitment Hour*, *Expendable* and, most recently, *Hunted* (Eos).



Ancient food for thought

Warwick Bray

Archaeological evidence of unexpected modes of food production in the tropics of lowland Central and South America carries lessons for modern farmers and students of crop-plant evolution.

In the debate about how to make rational, sustainable use of the American tropics, some of the more enlightened governments are taking a hard look at indigenous farming practices. Two papers (one on page 190 of this issue¹, the other in the 19 October issue of *Nature*²) show that archaeologists have a great deal to contribute by revealing crop regimes and forms of land management that have no counterparts today.

The conventional view that seasonally flooded savannas are worthless for cultivation and are suitable only for cattle ranching has been overturned by Erickson's work¹ in the Bolivian Amazon (Fig. 1). He has used aerial photographs and ground surveys to map a 'lost' prehispanic landscape of man-made earthworks, with settlement mounds, causeways, raised and drained fields, and also 500 km² of weirs and artificial ponds — which, he argues, were used for fish harvesting on a massive scale.

Erickson does not say whether the Bolivian government plans to reactivate the fish weirs, but experiments have already begun with his raised fields³. The archaeological sites yielded pollen of edible tubers (*Xanthosoma* species), and both maize and manioc (cassava) did well on the modern experimental plots. Several older farmers told Erickson it was the first time they had seen the pampas produce agricultural crops. Similar, and successful, experiments have been carried out with high-altitude raised fields in the Titicaca basin.

This may be just the start of things. In the savannas of Caribbean Colombia there are 5,000 km² of abandoned prehispanic fields and canals, unknown until the 1960s, and there are other, smaller expanses of raised fields in Mexico, Guatemala, Belize, Venezuela, Surinam, Guyana, Ecuador and Peru. Written documents from post-conquest times say nothing about the technology of these systems. But if further experiments show that the ancient agricultural landscapes can be reconstituted, and the savannas can be made productive, we may have an alternative to reckless felling of the rainforest.

Archaeological data of this kind help to resolve the dispute between historical anthropologists who accept European reports of large populations along the Amazon floodplain and in the savannas, and those members of the calorie-counting school who claim that protein scarcity



Figure 1 Central and northern South America today. Erickson's paper¹ describes investigations in the Bolivian Amazon and the identification of prehispanic raised fields, as well as weirs and artificial ponds. Piperno *et al.*² report evidence of starch grains 5,000–7,000 years old from the Aguadulce rock shelter, Panama. Taken together with other evidence, these findings point to the early and independent development of root-crop agriculture in the lowland American tropics.

imposed a limit on population growth in the lowland tropics until the introduction of European livestock. There is good evidence that a combination of game, fish, maize and palm fruits in seasonally flooded areas provides abundant protein⁴. Instead of worrying about protein, we should perhaps now think more about where the starch came from.

Taking the long archaeological view, this is not a silly question, especially for the initial stages in the evolution of agriculture. In the 1970s and 1980s archaeologists began to recover pollen, phytoliths (microscopic bodies of silica found in plant cells) and carbonized plant tissue from tropical sites and, more recently, to identify tuberous-plant starch grains that are not normally preserved. Piperno and colleagues² report the finding of starch grains from cultivated manioc, yams, arrowroot and maize on milling stones excavated at the Aguadulce rock shelter, Panama, and dated to between 5,000 and 7,000 years ago. This is the earliest evidence for root-crop cultivation in the Americas, and raises a series of questions.

The first generation of domesticated plants will look exactly like the last generation of wild ones, so botanists will never be able to recognize the initial experiments with cultivation. In Panama, the cultivated plants had already begun to diverge from their wild prototypes. The presence of maize and manioc is also significant, as these are not native to Panama and must have been brought into cultivation elsewhere. The genetic history of maize is still controversial⁵ but the consensus is that the botanical centre of origin is somewhere in Mesoamerica (Mexico and northern Central America), and most plausibly in Mexico. Manioc was probably domesticated in southwest Brazil⁶. These plants must therefore have reached Panama by a process of secondary dispersal, as domesticated crops and after a period of genetic adaptation to new ecological conditions.

The implication here is that we should be looking for a still earlier stage of tropical agriculture, without maize or manioc (the principal sources of starch today), and characterized by experimentation with whatever

was locally available. There are hints of this in the pollen and phytolith cores from the lake of La Yeguada, not far from the Aguadulce site⁷. Here, below levels with maize pollen dated to about 5,700 years ago, there is a sudden and large rise, around 11,000 years ago, of particulate charcoal and of plants typical of forest gaps. This phenomenon may well represent small-scale forest clearance for horticulture.

Like Erickson's fish farming, the earliest forms of tropical cultivation may be unlike anything that can be seen today. But recent excavations all over the lowlands point to the importance of tubers such as yams, arrowroot, sweet potato and *Xanthosoma*, and also emphasize the vital role of palm products. Arrowroot, a neglected crop today, is beginning to emerge as a significant early domesticate, with finds from highland Colombia dated to 10,000–9,000 years ago⁸ and from coastal Ecuador to around 3,600 years ago⁹. All this tends to confirm what Carl Sauer¹⁰ suggested as early as 1952: that there was an early and independent development of root-crop agriculture in the lowland American tropics, comparable in importance to the maize-beans-squash agriculture of Mesoamerica or the high Andean combination of potatoes, seed crops (quinoa) and llamas.

Perhaps we should now bury for ever the concept of 'centres of origin' for American agriculture. Unlike plants and animals, ideas are not constrained by ecological conditions. Archaeology is beginning to demonstrate the huge variety of agricultural practices in pre-European America, and also to suggest that people everywhere began by experimenting with the cultivation of plants they were already collecting in the wild.

Warwick Bray, formerly of the Institute of Archaeology, University College London, is at 29 Clarence Road, St Albans, Herts AL1 4NP, UK.

1. Erickson, C. L. *Nature* **408**, 190–193 (2000).
2. Piperno, D. R., Ranere, A. J., Holst, I. & Hansell, P. *Nature* **407**, 894–897 (2000).
3. Erickson, C. L. in *Archaeology in the American Tropics: Current Analytical Methods and Applications* (ed. Stahl, P.) 66–95 (Cambridge Univ. Press, 1995).
4. Beckerman, S. *Am. Anthropol.* **81**, 533–560 (1979).
5. MacNeish, R. S. & Eubanks, M. W. *Latin Am. Antiquity* **11**, 3–20 (2000).
6. Olsen, K. M. & Schall, B. A. *Proc. Natl Acad. Sci. USA* **96**, 5586–5598 (1999).
7. Piperno, D. R. in *Archaeology in the American Tropics: Current Analytical Methods and Applications* (ed. Stahl, P.) 130–153 (Cambridge Univ. Press, 1995).
8. Gnecco, C. & Mora, S. *Antiquity* **71**, 683–690 (1997).
9. Pearsall, D. M. in *Archaeology in the American Tropics: Current Analytical Methods and Applications* (ed. Stahl, P.) 126 (Cambridge Univ. Press, 1995).
10. Sauer, C. O. *Agricultural Origins and Dispersals* (Am. Geog. Soc., New York, 1952).

Surprisingly, low levels of glucocorticoids instead induced the release of MIF from macrophages³. MIF was then found to override or counter-regulate the suppressive effects of glucocorticoids on the expression of pro-inflammatory cytokines^{3,5}. This may account for many of MIF's pro-inflammatory effects *in vivo*^{3–5}.

MIF is generally released from pituitary cells, macrophages and T cells. The underlying mechanism remains largely a mystery, as does the process by which extracellular MIF exerts its effects on target cells. But perhaps the most unusual feature of MIF is its enzymatic activity. The crystal structure of MIF is strikingly similar to that of several small bacterial enzymes called isomerases⁸. Nevertheless, it seems doubtful (given the kinetic values calculated) that MIF interacts naturally with any of the substrates identified for it in biochemical studies. And mutation analyses have not supported the idea that MIF needs a working catalytic site for any of its tasks in the immune system^{8–10}.

MIF also seems to allow cells to bypass death (apoptosis) mediated by the p53 tumour-suppressor protein¹¹. So, because p53 function is often altered in tumours, MIF provides an important link between long-term inflammation and cancer. MIF is also involved in the sustained activation of cell-growth-promoting enzymes such as the ERK-1/2 family of mitogen-activated protein kinases¹², in promoting the growth of new blood vessels to nourish tumours¹³, and in regulating antitumour T cells¹⁴. A variety of results^{3–7,9–13} suggest that an interaction between MIF and a receptor on the surface of its target cells is essential for MIF's activities. But there remains scant information about what this receptor might be.

Kleemann *et al.*¹ now report an intriguing mechanism that bypasses the need for a cell-surface receptor (Fig. 1). They show that MIF binds to a cytoplasmic protein, Jab1. This protein usually induces both the phosphorylation of c-Jun, a protein involved in inducing cell growth, and the activity of AP-1, a transcription factor that activates the expression of pro-inflammatory genes. Jab1 also binds to and promotes the degradation of p27^{Kip1}, a protein that halts the cell-division cycle. By doing this, Jab1 can rescue serum-starved fibroblasts from growth arrest. MIF, however, inhibits Jab1, allowing levels of p27^{Kip1} to rise; when overexpressed, MIF reduces the growth-promoting effects of Jab1 on fibroblast cells (Fig. 1).

The physical interaction between MIF and Jab1 also suggests a way by which MIF, by inhibiting Jab1, might control the pro-inflammatory effects of AP-1 (Fig. 1). It seems that the cellular effects of MIF may now be interpreted as anti-inflammatory.

Signal transduction

A most interesting factor

Richard Bucala

The effects of the signalling molecule MIF are quite well understood, but how it works remains a mystery. Some of the pathways behind its activity have now been revealed — with surprising results.

The names — and acronyms — bestowed on proteins tend to reflect what they are doing at the time they are discovered. For example, one particular cytokine protein was found to inhibit the random migration of macrophages (immune cells that engulf foreign material) in culture. This led to it being called MIF, for 'macrophage migration inhibitory factor'. But this acronym also has another, less formal definition — 'most interesting factor'. Indeed, the functions of MIF, in the immune system and elsewhere, are diverse and intriguing. But the signal-transduction pathways underpinning MIF's activity have been unclear, so it has often been thought that this cytokine, although the first to be discovered, might turn out to be the last to be fully understood.

On page 211 of this issue, however¹, Kleemann and colleagues provide fresh insight into how MIF works, by showing that it interacts inside cells with an activator of gene expression, Jab1. The authors also propose a new, unconventional signalling mechanism

to account for MIF's interaction with this protein.

Although MIF was identified in the early 1960s, the gene encoding it was not cloned until 1989 (ref. 2). Christened for its effects on macrophages, MIF is still known as a key regulator of innate and acquired immunity — the two broad categories of immune response in higher organisms. For example, it is important in inducing inflammation in response to invasion by bacteria, viruses and so on. But its biochemical and biological properties are not at all predictable^{3,4}.

For example, MIF is involved in regulating the activation of macrophages and T cells^{3,5}, in the release of insulin from the pancreas⁶, and in carbohydrate metabolism⁷. The mouse counterpart of MIF is released by the anterior pituitary gland as a consequence of the systemic response to stresses^{3,4}.

The informal name for MIF was coined by Thierry Calandra after an attempt to block its expression by using glucocorticoids — steroid hormones thought to suppress the expression of pro-inflammatory cytokines.

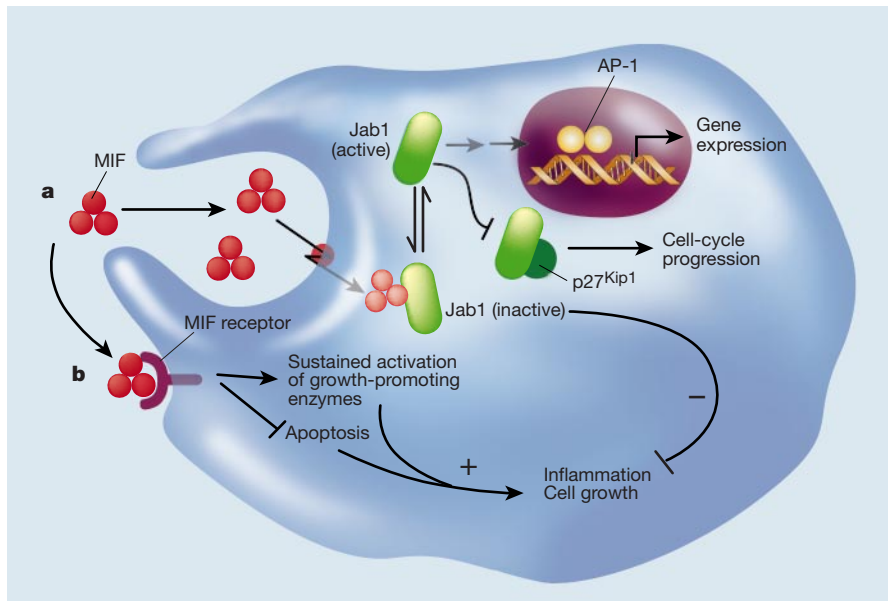


Figure 1 Pathways by which MIF (macrophage migration inhibitory factor) may regulate inflammation and the proliferation of target cells. **a**, Kleemann *et al.*¹ suggest that MIF can be taken up into target cells by endocytosis. This probably occurs when MIF is present at high extracellular concentrations. Once in the cytoplasm, MIF interacts with Jab1, inhibiting it. In the absence of MIF, Jab1 activates a transcription factor, AP-1, that activates pro-inflammatory genes; Jab1 also leads to the degradation of a cell-cycle inhibitor, p27^{Kip1}. In the presence of MIF, however, Jab1 cannot exert these effects, leading to negative effects on inflammation and cell growth. **b**, At low concentrations, MIF may bind to a specific receptor, leading directly to intracellular signals that have a positive effect on inflammation and cell growth.

This might seem odd given the known pro-inflammatory effects of MIF. And MIF's growth-inhibitory effects, through enhanced p27^{Kip1} expression, also contrast with its cell-growth-promoting and cell-death-suppressing actions in other cell systems^{11,12}. However, one feature of MIF is its bell-shaped dose-response curve with respect to several biological phenomena, such as its inhibition of macrophage migration. This means that high MIF concentrations may actually inhibit several biological processes.

Kleemann *et al.*¹ propose that cells take up MIF by endocytosis—a process by which a cell's outer membrane engulfs extracellular substances (often in a nonspecific way) and imports them into the cell in membrane-bound vesicles. MIF could then meet up with Jab1 in the cytoplasm (Fig. 1). But how does MIF—or its active domain—move out of the vesicle? That problem aside, this model seems plausible, and would require high extracellular concentrations of MIF. However, extracellular MIF levels range from about 4 ng ml⁻¹ in normal blood plasma to 150 ng ml⁻¹ in patients with septic shock¹⁵. How can this model explain the established pro-inflammatory and growth-promoting effects of MIF at the lower end of this range? At these concentrations, one would expect MIF to need to interact—with high affinity—with a specific receptor on the cell surface that would then send signals into the cell.

It seems likely that MIF does indeed regulate the activity of Jab1; Jab1 may even have a direct role in regulating MIF expression. But it may be premature to conclude that these results represent a breakthrough in our understanding of how MIF works. Nonetheless, the ultimate value of a model lies in its usefulness for predicting normal physiology and disease. The unusual mechanisms of uptake and activity now proposed for MIF by Kleemann and colleagues¹ certainly make for several intriguing predictions. They are likely to ensure that MIF retains its reputation as a most interesting factor.

Richard Bucala is at The Picower Institute for Medical Research, 350 Community Drive, Manhasset, New York 11030, USA.
e-mail: rbucala@picower.edu

1. Kleemann, R. *et al.* *Nature* **408**, 211–216 (2000).
2. Weiser, W. Y. *et al.* *Proc. Natl Acad. Sci. USA* **86**, 7522–7526 (1989).
3. Calandra, T. *et al.* *Nature* **377**, 68–71 (1995).
4. Bernhagen, J. *et al.* *Nature* **365**, 756–759 (1993).
5. Bacher, M. *et al.* *Proc. Natl Acad. Sci. USA* **93**, 7849–7854 (1996).
6. Waerber, G. *et al.* *Proc. Natl Acad. Sci. USA* **94**, 4782–4787 (1997).
7. Benigni, F. *et al.* *J. Clin. Invest.* (in the press).
8. Swope, M. D. & Lolis, E. *Rev. Physiol. Biochem. Pharmacol.* **139**, 1–32 (1999).
9. Bendrat, K. *et al.* *Biochemistry* **36**, 15356–15362 (1997).
10. Hermanowski-Vosatka, A. *et al.* *Biochemistry* **38**, 12841–12849 (1999).
11. Hudson, J. D. *et al.* *J. Exp. Med.* **190**, 1375–1382 (1999).
12. Mitchell, R. A. *et al.* *J. Biol. Chem.* **274**, 18100–18106 (1999).
13. Chesney, J. *et al.* *Mol. Med.* **5**, 181–191 (1999).
14. Abe, R. *et al.* *J. Immunol.* (in the press).
15. Calandra, T. *et al.* *Nature Med.* **6**, 164–170 (2000).



100 YEARS AGO

At the request of the Austrian Ministry of Agriculture, various experiments have been made by Drs Pernter and Trabert with the view of testing the use of Mr Stiger's apparatus for dispersing hail-clouds by gun-firing. The apparatus consists of a mortar with a long funnel fixed to the orifice; upon firing a sufficient charge of powder, rings or whirls are formed in the air and can be followed either by their hissing sound or by the particles of smoke carried up with them. The force and durability of the whirls vary with the charge, and with the size of the funnel, but it does not appear from the experiments that a greater altitude than about 400 metres was reached, which is much less than had been previously stated. It does not seem probable, therefore, that unless the hail-clouds are very low that any practical result is likely to be attained. The most that can be said in favour of the process is that while in some cases the formation of hail may have been prevented by the disturbance of equilibrium, hail frequently falls, in spite of frequent firing.

From *Nature* 8 November 1900.

50 YEARS AGO

The Nobel Prize for Medicine for 1950 has been awarded jointly to Dr. Philip Showalter Hench, of the Mayo Clinic, Rochester, Minnesota, Prof. Edward Calvin Kendall, of the Mayo Foundation, Rochester, and Prof. Tadeus Reichstein, of the University of Basel. The award can be regarded as a recognition of the work which has been done leading to the dramatic opening of a new field of scientific medical research—that of rheumatoid arthritis and related diseases. Early in 1949, Hench, Kendall, Slocumb and Polley (*Proc. Staff Meet. Mayo Clinic*, **24**, 181; 1949) reported on “the effects of the hormone of the adrenal cortex (...Compound E) and of pituitary adrenocorticotrophic hormone on rheumatoid arthritis”. This was followed soon after by a report of the effects of compound E on rheumatic fever (Hench, Slocumb, Barnes, Smith, Polley and Kendall, *Proc. Staff Meet. Mayo Clinic*, **24**, 277; 1949). Since then, investigations have begun in centres of research in all parts of the world on the physiological action and the clinical application of cortisone. This work is now developing at an increasing rate, limited only by the availability of cortisone.

From *Nature* 11 November 1950.

Aquatic ecology

A lake's life is not its own

George W. Kling

Little is known about how lakes change as they get older. Research on a 12,000-year sequence of natural laboratories at Glacier Bay, Alaska, shows how the terrestrial surroundings can influence lake evolution.

Imagine a landscape suddenly released from the frozen grip of a kilometre-thick glacier. Bare rock crumbles to pockets of dust that harbour sprouting seeds, and soon patches of plant species adapted to this environment appear. As the environment gradually changes, the ensuing development of the flora — a process known as succession — proceeds roughly in step with predictions from theories of plant community dynamics. As ecology goes, we have a good idea of what happens in these circumstances.

But what of the numerous lakes left in depressions after the glacier has retreated? Surely science can also predict their development? Well... no. Unlike terrestrial ecology, which is founded in part on numerous studies of plant succession, aquatic ecologists have only the sketchiest of ideas about succession in lakes; there is no workable theory of how their chemistry and biology evolve. On page 161 of this issue¹, Engstrom *et al.* present a powerful data set that provides insights into what happens in the first few years of a lake's life — and also the following thousands of years. More importantly, they can give reasons why the lakes they studied evolved as they did.

Why aren't there any general theories of lake development? One of the original ideas from the founder of this field, the late Ed Deevey at Yale University, was that lakes become more productive as they age; that is, the amount of biomass created in the lake increases. Although that is sometimes true, and the notion is often taken as accepted fact, it has long since been evident that it simply

fails to hold as a generality². There are also studies of how lakes develop towards the end of their existence, such as Walker's³ classic analysis from the British Isles of how lakes become land. But the variable patterns of succession from open water to swamps, and from bogs to forest, were predictable only in the coarsest statistical sense.

Perhaps the big problem is that the chance of observing lake evolution is so rare. Yes, there are hundreds of artificial reservoirs that we could track through history, but most of them flush so freely that they never truly develop as lakes, surviving instead as wide, slowly flowing rivers. And the approach of comparing lakes formed at different times is difficult, because most lakes were formed at the end of the last glaciation, at least 10,000 years ago. That makes the gap between recent reservoirs and natural lakes too great, and interpolation between those two points forms an uninteresting line.

Enter Engstrom *et al.*¹, who have taken advantage of a natural experiment — the advance and final retreat of ice at Glacier Bay, Alaska, over the past 12,000 years. The end result is a superb sequence of natural laboratories in the form of lakes ranging from 10 years to 12,000 years in age (see the map on page 162 of the paper). The authors looked first across all the lakes for patterns in diatom species (lake algae) and water chemistry that were related to the age of the lake. They next studied the history of each lake, as recorded in the form of fossil diatoms and the chemistry of sediments buried over time. The patterns they found in the results of these

comparative and historical approaches match, which is wonderful as it adds great confidence to the group's conclusions.

Engstrom and colleagues show that their lakes have become more dilute, more acid and more coloured with dissolved organic matter over time (and, by inference, less rather than more productive). The authors' explanations are intricate but don't hide the main picture that the terrestrial surroundings have controlled lake evolution. But this control is not as simple as the depletion of easily weathered minerals in the catchment area causing the runoff waters to become progressively more dilute. Rather, there appears to be a 200-year lag in the initial changes in lake chemistry and biology, a lag which coincides with the development of soils and vegetation in the Glacier Bay area. Such soil development eventually excludes water from running through the deeper, mineral-rich substrates and into the lakes. Even the details of variation between individual lakes in Engstrom and colleagues' study can be explained by the geomorphological control of local hydrology.

These results¹ will help to reshape our understanding of lake evolution. The eventual goal is to predict how aquatic ecosystems respond to the interaction of landscape geomorphology, climate change and ecological factors acting through terrestrial vegetation or within the lake itself. This is the real mystery story, with the potential factors and mechanisms being so intertwined that it seems hard to isolate a problem let alone an answer.

Yet there is hope. We now have Engstrom and colleagues' evidence that geomorphological controls can dominate lake development in some cases. And Birks *et al.*⁴ have shown that sudden variations in aquatic ecosystems resulted from rapid climate change (during the last glacial-to-interglacial transition, about 14,000 years ago), while longer-term variation since then resulted from the interplay of climate and changes in the catchment area and in the lakes themselves.

In a further twist⁵, it seems that a lake's response to a strong shift in climate may be overridden by geomorphological controls. For example, arctic Alaska experienced a regional change in climate about 7,000 years ago⁶. This resulted in dramatic alterations in the carbon chemistry of sediments in a lake on a young, topographically varied landscape, but little or no stratigraphic variation in a neighbouring lake on a much older, smoother landscape (Fig. 1). The two lakes were exposed to identical climates and are similar in nature. Yet these indicators of carbon inputs or lake metabolism varied greatly across the climate transition in only one of the lakes.

In revising our ideas about lake development, it is clear that climate, geomorphology and ecology are all major players that can smother each other now and then. We must discover patterns of when and where this has

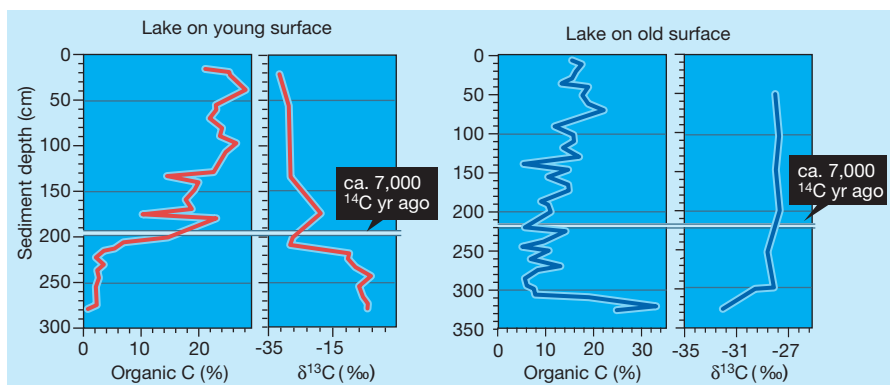


Figure 1 Variation of organic-carbon content and $\delta^{13}\text{C}$ value in the sediments of two lakes in arctic Alaska⁵. Both lakes experienced the same climatic transition about 7,000 years ago. But the carbon chemistry (an indicator of processes happening in the catchment area or the lake itself) responded only in the lake on the young surface. This example shows how landscape characteristics can override the effects on lake development from even rapid and strong climate change.

happened in the past, as these will guide our thinking about how aquatic ecosystems will change in the future. This great paper¹ pushes us along, and it'll be fun to see how it all works out.

George W. Kling is in the Department of Biology, University of Michigan, Ann Arbor, Michigan 48109-1048, USA.

e-mail: gwk@umich.edu

1. Engstrom, D. R., Fritz, S. C., Almendinger, J. E. & Juggins, S. *Nature* **408**, 161–166 (2000).

2. Livingstone, D. A. *Am. J. Sci.* **255**, 364–373 (1957).
3. Walker, D. in *Studies in the Vegetational History of the British Isles* (eds Walker, D. & West, R. G.) 117–139 (Cambridge Univ. Press, 1970).
4. Birks, H. H., Batterbee, R. W. & Birks, H. J. *J. Paleolimnol.* **23**, 91–114 (2000).
5. Hu, F. S., Oswald, W. W., Brubaker, L. B. & Kling, G. W. in *Geomorphic and Climatic Controls over Holocene Soil Development on the Arctic Foothills of Alaska 204* (Ecol. Soc. Am. 85th Annual Meeting, 6–10 August; Snowbird, Utah, 2000).
6. Brubaker, L. B., Anderson, P. M. & Hu, F. S. in *Arctic and Alpine Biodiversity* (eds Chapin, F. S. & Körner, C.) 111–125 (Springer, Berlin, 1995).

come the silencing response is to move rapidly throughout the plant before silencing can shut them down. Such systemic spread of the virus requires virus-encoded movement proteins. These come in a variety of forms⁵: some are enzymes, while others have structural roles. Some of them localize to plasmodesmata (the channels connecting plant cells), affect the ability of these channels to open and close, and stimulate transport of themselves and nucleic acids (the viral genome) to the neighbouring cell. In many cases, the functions of the movement protein are split among several proteins. For example, the RNA virus potato virus X (PVX) has three movement proteins — p25, p12 and p8. The conventional view is that these proteins coordinate the movement of the virus to plasmodesmata, the alteration of the channels to accommodate the virus, and transit of the virus through the channels.

This model has merit but — given the finding by Voinnet *et al.*¹ that p25 is a suppressor of systemic RNA silencing — it may be too limited. The p25 protein is not the first silencing suppressor in a plant virus to be identified^{6–8}. But the fact that it is also a movement protein indicates that viral movement might depend both on interactions between movement proteins and plasmodesmata, and on suppression of the plant's silencing mechanism.

Two of Voinnet *et al.*'s results¹ are worth mentioning in more detail here. First, p25 specifically inhibits the production of the systemic silencing signal. This was shown using plants engineered to express a transgene

RNA silencing

Moving targets

James C. Carrington

Viruses have evolved several strategies to attack plants, but the plants keep hitting back. So the viruses have upped the ante by stopping the plants' immune response from spreading to uninfected tissues.

Plants have developed a variety of weapons in their battle against viral invaders, one of the most elegant of which is 'RNA silencing'. The beauty of this immune response lies in its ability to adapt to all sorts of different viruses, because its specificity is dictated by the sequence of the viral genome itself. Another clever feature is that a signal that triggers silencing is not restricted to individual plant cells, but can spread from the site of infection, generating a response in more distant tissues. But it seems that viruses do not admit defeat easily. Both viruses and these plant-wide ('systemic') silencing signals spread from cell to cell through channels that traverse the cell walls. The spread of the viruses is dependent on viral 'movement proteins'. Writing in *Cell*¹, Voinnet and colleagues show that these movement proteins may also debilitate the systemic arm of the silencing response.

RNA silencing can be triggered in plants by replicating viruses, double-stranded RNA molecules and foreign genes (transgenes) that allow the production of high levels of normal or aberrant messenger RNAs². Silencing induced in this way by viruses limits the accumulation of the inducing virus, and can also confer immunity in the plant against closely related viruses.

The best model for how RNA silencing works³ suggests that double-stranded segments within an RNA target, such as a viral RNA, are recognized by a plant RNA-digesting enzyme (a 'dsRNase') that contains regions for binding to and untwisting double-stranded RNA. This nuclease cleaves the target RNA to produce segments 21–23 nucleotides in length. These small RNAs associate with the dsRNase or another RNA-degrading nuclease, and confer specificity in further RNA-degradation reactions because they can pair up with complementary single-stranded target RNA.

Viral genomes come in DNA or RNA forms; viruses with RNA genomes may be strong inducers of RNA silencing because double-stranded RNA is formed, at least temporarily, during replication of the genome. The silencing mechanism may halt the replication of the virus in the infected area. Moreover, plant tissues far from the infected site also show specificity in silencing the viral RNAs⁴. The most plausible explanation for this is that a mobile signal — containing nucleic acids corresponding to the RNA target — spreads from cell to cell.

One strategy that viruses deploy to over-

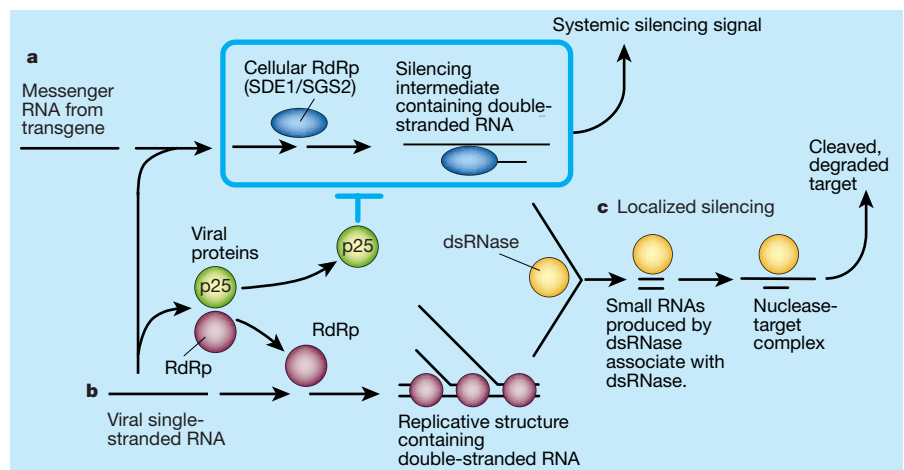


Figure 1 Plants may have a two-pronged strategy for silencing viruses, as suggested by Voinnet *et al.*¹. a, The 'weak-inducer' branch. RNAs with limited double-stranded structure — such as those encoded by foreign genes (transgenes) or resulting from replication of a viral genome — feed into this pathway. Silencing depends on a cellular RNA-dependent RNA polymerase (RdRp)-like protein (such as SDE1/SGS2). SDE1/SGS2 produces double-stranded RNA using the single-stranded RNA as a template. This branch also results in the production of a systemic silencing signal and is inhibited by a movement protein, p25, from the potato virus X. b, The 'strong-inducer' branch. Replication of the genome of an RNA virus, using a viral RdRp, produces some replication intermediates with double-stranded-RNA structure. This pathway is not inhibited by p25 and does not lead to systemic silencing. c, Both pathways lead to localized silencing. A 'dsRNase' enzyme cuts up the double-stranded RNAs. The resulting small RNAs associate with the dsRNase or another RNA-degrading nuclease, which cleaves target RNAs that pair up with the small RNAs.

encoding green fluorescent protein (GFP). GFP is a convenient molecule for these sorts of studies because if the GFP transgene is silenced the plants do not glow bright green under ultraviolet light. In Voinnet *et al.*'s study, silencing of this transgene — both locally and plant-wide — was induced by localized injection of either a PVX–GFP 'recombinant' virus or a highly expressed GFP gene. But when p25 was injected into the lower leaves of the plants simultaneously with either inducer, silencing was blocked in the upper leaves of the plants.

Second, although p25 arrested systemic silencing by either inducer, it inhibited localized silencing induced by the injected GFP gene but not by the replicating PVX–GFP recombinant virus. This surprising result is explained if there are two avenues by which localized silencing can be triggered, only one of which leads to systemic silencing (Fig. 1). One branch is activated by both the injected GFP gene (or rather, its mRNA) and the PVX–GFP viral RNA; this branch is necessary for production of the systemic signal, and is inhibited by p25. Given that both local and systemic silencing are arrested when the injected GFP gene is used together with p25, it seems likely that p25 inhibits this pathway before the silencing signal is even produced. The other branch results only in local silencing; it is activated only by replicating PVX–GFP virus, is not suppressed by p25, and does not lead to production of the systemic signal. So, for replicating PVX–GFP virus, systemic silencing occurs through the p25-sensitive branch, but local silencing occurs through the p25-insensitive pathway.

Why do these two pathways exist? It is likely that there are differences in the silencing-inducing RNAs formed by the GFP gene and the replicating virus. Expression of the injected GFP gene yields a functional mRNA that is predicted to adopt limited double-stranded RNA structure — such molecules are probably 'weak' inducers of silencing³. Such weak inducers are predicted to be processed through the p25-sensitive, systemic-signal-producing pathway. This pathway may require a cellular RNA-dependent RNA-polymerase-like enzyme, such as SDE1/SGS2, which is necessary for RNA silencing induced by transgenes but not by several replicating viruses^{9,10}. Replication of the PVX–GFP virus, on the other hand, yields both single-stranded RNA with limited double-stranded structure, and replication intermediates with more double-stranded RNA structure. The single-stranded viral RNA may enter the 'weak-inducer' pathway. In contrast, the double-stranded RNA formed during replication might be a 'strong' inducer that serves directly as a substrate for the dsRNase in a p25-insensitive manner.

The details of most key steps in this systemic signalling pathway remain to be determined. But one needs to look no further than

the viruses to find new entry points into this fast-moving field.

James C. Carrington is at the Institute of Biological Chemistry, Washington State University, Pullman, Washington 99164-6340, USA.
e-mail: carrington@wsu.edu

1. Voinnet, O., Lederer, C. & Baulcombe, D. C. *Cell* **103**, 157–167 (2000).

2. Meins, F. *Plant Mol. Biol.* **43**, 261–273 (2000).

3. Bass, B. L. *Cell* **101**, 235–238 (2000).

4. Fagard, M. & Vaucheret, H. *Plant Mol. Biol.* **43**, 285–293 (2000).

5. Lazarowitz, S. G. & Beachy, R. N. *Plant Cell* **11**, 535–548 (1999).

6. Anandalakshmi, R. *et al. Proc. Natl Acad. Sci. USA* **95**, 13079–13084 (1998).

7. Brigneti, G. *et al. EMBO J.* **17**, 6739–6746 (1998).

8. Kasschau, K. D. & Carrington, J. C. *Cell* **95**, 461–470 (1998).

9. Dalmay, T. *et al. Cell* **101**, 543–553 (2000).

10. Mourrain, P. *et al. Cell* **101**, 533–542 (2000).

Nanotechnology

In control of molecular motion

Ben L. Feringa

Nature leads the way when it comes to motors on a molecular scale. But chemists are in hot pursuit, designing controllable structures that can mimic muscles or rotary motors.

If you could build a motor one millionth of a millimetre across, you could fit a billion billion of them on a teaspoon. It seems incredible, but biological systems already use molecular motors on this scale. Linear motion is produced by skeletal muscle¹, and rotary motion by the enzyme ATP synthase². The underlying principle of these natural motors is the production of continuous motion while energy is being consumed. Can scientists engineer similar movements in a synthetic system?

Two recent findings suggest a way of using chemical synthesis to build molecular-scale components for artificial motors. In *Angewandte Chemie International Edition*, Sauvage and co-workers³ report a molecular assembly that can contract and stretch as desired under the influence of a chemical signal. It is the first synthetic molecule in which two filaments

can glide along each other, as they do in real muscles. And in the *Journal of the American Chemical Society*, Aida and co-workers⁴ describe the acceleration or deceleration of motion in rotating disc-like structures in response to changes in their electronic state. By using a metal ion sandwiched between two porphyrin ligands, the rotary motion can be controlled by oxidation (removing electrons) or reduction (adding electrons) at the metal centre, producing a device reminiscent of transmission systems in much larger motors.

Most artificial systems that produce linear or rotary motion are more like molecular switches than motors. It is not enough to build a system that moves or rotates freely. To reproduce the actions of a motor it is essential to incorporate control into the system, so the motion can be started and stopped at will. There are many ways to do this. For example,

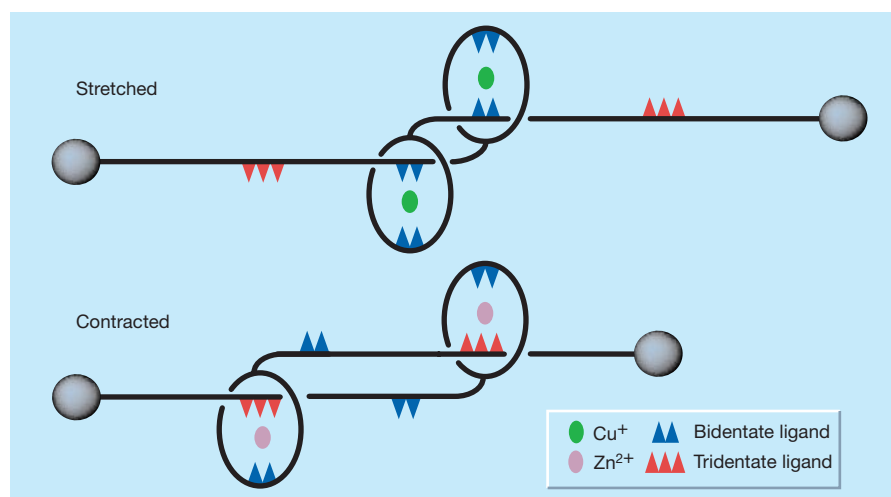


Figure 1 Stretching and contracting in an artificial molecular muscle designed by Sauvage and co-workers³. Two units, each containing a string and a ring, are threaded together, and prevented from coming apart during movement by bulky stoppers (grey circles) at the end of the strings. Two metal ions can simultaneously bind to the ring and the strings. Copper ions (green) prefer binding to two bidentate ligands (blue), whereas zinc ions (pink) prefer binding to a bidentate and a tridentate ligand (blue and red). The stretched configuration occurs when two copper ions are bound. When zinc ions replace the copper ions, the strings slide along each other as the metal ions seek out their most favourable binding position. As a result the artificial muscle contracts.

polyacrylamide gels⁵ and polypyrroles⁶ have been designed to mimic the behaviour of muscles by bending or contracting in response to electric current. Other approaches to linear motion rely on the use of rotaxanes — molecular rings threaded by strings — in which the position of the ring can be altered by electrochemical or photochemical means^{7,8}.

In their system, Sauvage and colleagues³ use a doubly threaded structure to mimic the stretching and contracting of a real muscle (Fig. 1). Their design is based on two molecular units, each consisting of a ring attached to a string, which can slide along each other. Each ring has a metal ion at its core, and there is a bulky stopper at the other end of the string to prevent the strings from coming apart. Each ring contains a bidentate ligand (so called because it has two points of attachment for a metal ion), whereas each string contains both a bidentate and a tridentate ligand (which has three points of attachment). The geometry of the structure changes, depending on which ligands bind to the metal ions.

In the presence of two copper ions, the most favourable geometry has two bidentate ligands binding to each ion. This is the 'stretched' geometry. Replacing the copper ions with zinc ions causes the strings to slide along each other so the zinc ions can adopt

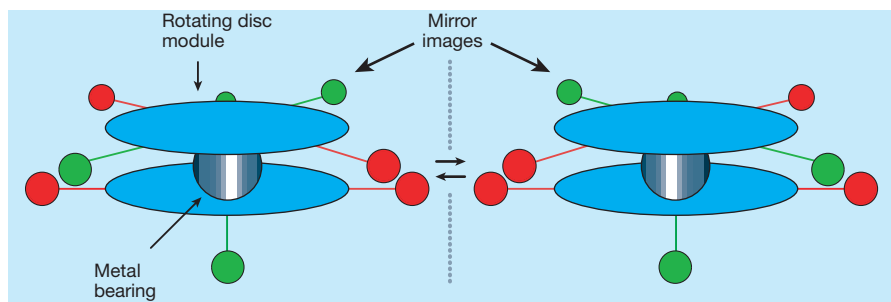


Figure 2 Control of rotation in a disc-like structure. In the molecular system created by Aida and co-workers⁴, cerium or zirconium ions (blue) are sandwiched between two porphyrin ligands. The metal ion acts as a molecular ball-bearing between two discs as they rotate. The side chains (green and red balls) are attached in such a way that they can form two mirror images, which change as the discs rotate with respect to one another, allowing the rate of rotation to be observed. Adding an electron to the cerium ion increases the distance between the rotating discs, weakening the interaction between the porphyrin discs, and resulting in faster rotation. Removing an electron from the zirconium ion reduces the distance between the discs, leading to slower rotation.

their preferred configuration, with each ion bound to one bidentate ligand and one tridentate ligand. This is the 'contracted' geometry. Replacing zinc with copper reverses the process, stretching the molecule. The length of the structure changes by roughly 27% — coincidentally, about the same as that seen in natural muscles.

How similar this is to the motion of muscle filaments is arguable. But this is the first

demonstration of stretching and contracting motions induced by a chemical reaction in an artificial molecular assembly. The stage is set for the design of other 'molecular muscles' triggered by electrochemical or photochemical processes. For these artificial molecular muscles to perform useful mechanical work, they will ultimately have to be connected to macroscopic systems. The organization and concerted action of a large number of these

Oceanography

Flow charts

Space-borne instruments have revolutionized research on the circulation patterns and strengths of the oceans. For instance, a satellite can observe all of the world's oceans in less than ten days. With radar technology, it can measure the shape of the sea surface and provide observations on the surface flows, including major currents or eddies — phenomena that are much like weather systems in the atmosphere.

The TOPEX/POSEIDON mission,

a US–French undertaking, has been prominent in providing precision observations of this sort. But a single satellite has its limitations in detecting variations both in space and time. Although there has been more than one radar instrument in space for several years now, the task of merging satellite data is formidable and requires considerable statistical skill. Yet that is what N. Ducet and colleagues have done, as they describe in the *Journal of*

Geophysical Research (105, 19477–19498; 2000). The result is the best picture yet of the ocean and its vigorous eddy component.

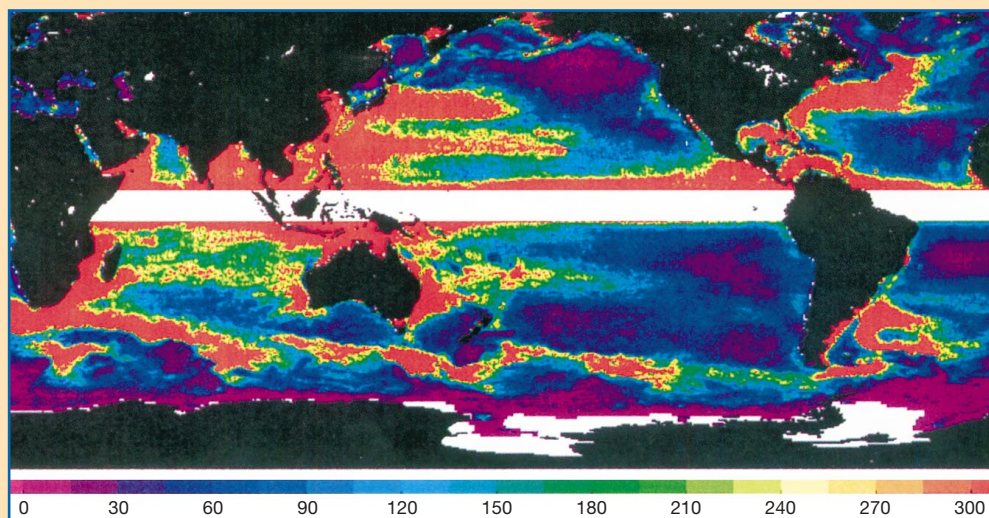
The authors have combined data from TOPEX/POSEIDON and the European ERS missions over the period 1992–98. The picture here shows in stunning detail the regions where currents fluctuate vigorously (units are in $\text{cm}^2 \text{s}^{-2}$). The regions of high (red) amplitudes indicate particularly turbulent currents that

are involved in transporting and redistributing heat and other properties, and so affect global climate.

What of the future? The next aim is to fly satellites routinely in a configuration and with technology that will allow current and eddy flows to be measured directly without the need for statistical estimation. The ultimate objective is to incorporate the data into models of ocean circulation, much as is done in weather forecasting. Physical oceanographers will not be the only beneficiaries. Other prospects include identifying local variations in flows that affect fisheries, the distribution of pollutants, and fuel-saving routes for ships. On a larger scale, such studies should also tell us more about the transport of heat from low latitudes towards the poles, where it is released to the atmosphere and has a large influence on mid-latitude climate.

Detlef Stammer

Detlef Stammer is in the Physical Oceanography Research Division, Scripps Institution of Oceanography, 9500 Gilman Drive, La Jolla, California 92093-0230, USA. e-mail: detlef@jford.ucsd.edu



molecular motors is one of the key issues to address next.

So much for linear motion, but how about rotary motion? Particularly promising are studies with interlocked rings (catenanes) in which circular motion can be induced^{7,8}. Last year, two studies demonstrated the controlled conversion of energy into unidirectional rotation, a fundamental property of a rotary motor^{9,10}. The next challenge is to control the rate of rotation to produce a motor capable of more than one speed.

Aida and co-workers⁴ have found an elegant solution to this problem with a molecular system called a bisporphyrinate double-decker complex (Fig. 2). In these complexes a cerium or zirconium ion is sandwiched between two porphyrin ligands that rotate with respect to one another. The metal ion functions as a kind of ball-bearing between two rotating discs. The configuration of bulky side chains attached to the porphyrin ligands means that the metal complexes are chiral — that is, they can be mirror images of each other. This feature allows the authors to study the dynamics of the rotary motion by measuring their optical activity. (Mirror-image chiral molecules rotate polarized light in opposite directions.)

Aida and colleagues⁴ found that reducing the cerium complex led to the rotation of the porphyrin ligand being accelerated more than 300-fold. Oxidation of the zirconium complex, on the other hand, decelerated it by a factor of 21 or 99, depending on the oxidation state of the complex. The change in rotary motion is attributed to a change in distance between the two porphyrin ligands. In the cerium complex, reduction of the metal centre increases the ionic radius, so the interaction between the porphyrin ligands weakens, which leads to faster rotation. In the zirconium complex, oxidation reduces the distance between the porphyrin ligands, strengthening the interaction between the ligands and slowing down the rotary motion.

The chemical control of rotation is a powerful tool to be used in more advanced molecular motors. But there are many hurdles to overcome before the structures designed by these and other groups lead to molecular machinery becoming a reality. To make their construction easier, self-assembly processes are needed. It is also not clear whether the motors described here will retain their properties when they are used as part of a more complex system. Nonetheless, we are adding important components to our toolkit for making nanomachines.

Ben L. Feringa is in the Department of Organic Chemistry, University of Groningen, Nijenborgh 4, 4947 AG Groningen, The Netherlands.

e-mail: feringa@chem.rug.nl

1. Abrahams, J. P., Leslie, A. G. W., Lutter, R. & Walker, J. E. *Nature* **370**, 621–628 (1994).

2. Rayment, I. *et al. Science* **261**, 58–65 (1993).

3. Jimenez, M. C., Dietrich-Buchecker, C. & Sauvage, J.-P. *Angew. Chem. Int. Edn* **39**, 3284–3286 (2000).

4. Tashiro, K. *et al. J. Am. Chem. Soc.* **122**, 7921–7926 (2000).

5. Li, Y. & Tanaka, T. *Annu. Rev. Mat. Sci.* **22**, 243–277 (1992).

6. Otero, T. F. & Sansinera, J. M. *Adv. Mater.* **10**, 491–494 (1998).

7. Balzani, V. *et al. Acc. Chem. Res.* **31**, 405–414 (1998).

8. Sauvage, J.-P. *Acc. Chem. Res.* **31**, 611–619 (1998).

9. Koumura, N. *et al. Nature* **401**, 152–156 (1999).

10. Kelly, T. R., De Silva, H. & Silva, R. A. *Nature* **401**, 150–152 (1999).

Computational neuroscience

Intimate attention

Jochen Braun

If we are to perceive a visual figure, we need to direct our attention towards it. The more we discover about this ability, the more impressive it seems.

Looking at Edgar Rubin's famous image of a white vase on a dark background — or is it two dark faces on a white background? — we can alter our perception by directing our attention to either vase or face (Fig. 1). The neural basis for this ability to perceive at will some parts of a visual scene as 'figure' and others as 'background' is a hotly debated area of research. On page 196 of this issue¹, Blaser and colleagues show that, by directing our attention in this way, we can distinguish even between two intimately associated 'figures' with almost identical characteristics. Their results have profound implications for how figures are represented in the brain's visual cortex.

The context of this work¹ is a debate over whether 'attention' is free to select arbitrary visual locations and attributes (any location, colour or shape of interest to the observer) as the figure, or whether it must select 'visual objects'^{2,3}. Here, a visual object is defined as a cluster of locations and attributes that are linked by the Gestalt rules of visual 'completion'. What these rules mean for the visual cortex is that neuronal responses to a given visual stimulus often depend on the context — that is, on other nearby stimuli⁴. There are strong arguments that the neuronal architecture of the visual cortex must limit the selective capabilities of attention^{5–8}. This is because there do not seem to be enough feedback connections to the early visual cortex to allow attention to affect arbitrary combinations of neurons.

To study how an observer divides a visual scene into 'figure' and 'background', Blaser *et al.*¹ build on their earlier work and use 'attentional tracking'^{9,10}. This stratagem, which pushes attentional capabilities to the limit, involves a dynamic visual display in which two or more possible figures change beyond recognition several times. If observers can maintain attention on one particular figure (that is, 'track' it successfully while it changes), they can say which part of the final display is descended from the initial figure.

Blaser *et al.* have now constructed 'mutable objects' with three independent attributes — orientation, width of stripes, and colour saturation — that change rapidly and continually, each following an independent sched-

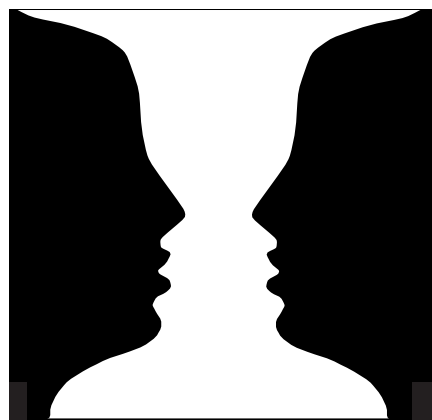


Figure 1 Is it a vase or two faces? In Rubin's classic illustration, attention can select either object as the 'figure', thereby altering perception.

ule (Fig. 2). The objects sometimes turn clockwise and sometimes anticlockwise. Stripe width sometimes increases and sometimes decreases. And the degree of colour saturation likewise sometimes increases (towards stripes of red on black) and sometimes decreases (towards stripes of light grey on black).

The authors superimposed two of these mutable objects (Fig. 2). Observers were asked to focus attention on one object, specified by its orientation, and to maintain attention on this object for ten seconds, during which time both objects changed. The impression of the observers was that, most of the time, the two objects seemed to remain perceptually distinct. At the end of the ten-second period, observers were able to report reliably which object — again specified by orientation — had descended from the initial object. In other words, observers were able to select one object as the figure, and to relegate the other to the background.

This ability to distinguish between objects found at the same location, and assuming the same attributes at different times, reveals hitherto unsuspected capabilities of visual attention. But what does attention select as the figure — the tracked object as a whole, comprising its orientation, stripe width and colour saturation? Or does it merely select the task-relevant attribute — in this case, orientation?

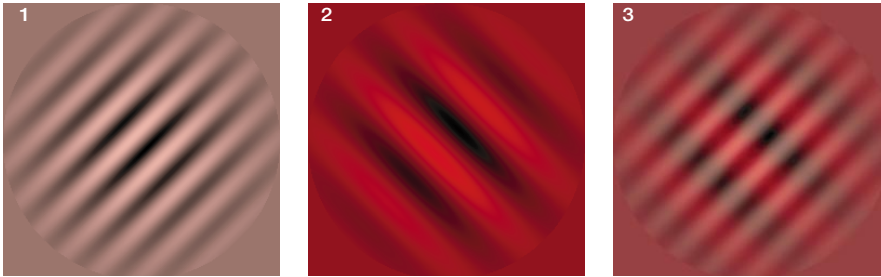


Figure 2 What does attention select as 'figure' when two objects have almost identical characteristics? To address this question, Blaser *et al.*¹ superimposed two objects, each with three attributes — orientation, stripe width, and degree of colour saturation. In the experiment, all attributes changed rapidly and continuously over time. In the snapshot shown here, object 1 has narrow stripes that are pink on black and orientated at +45°, whereas object 2 has broad stripes that are red on black and orientated at -45°. The objects were superimposed (3) and observers were asked to maintain attention on one of the objects as the attributes changed. Even after several seconds, observers were able to state which object in the changed display had descended from the initial object. Further results suggest that attention may select all three attributes of one object as 'figure', but not a single attribute or a mixture of attributes from both objects.

To answer this question, Blaser *et al.* used a modified display of two mutable objects and asked observers to monitor either one or two attributes of the tracked object. During the tracking period, all attributes of both objects showed simultaneous, discontinuous jumps, which observers had to report. Observers were able to monitor any two attributes of the same object about as well as they monitored either attribute alone. The implication is that all attributes of the tracked object are perceived as 'figure'. (If only task-relevant attributes were perceived as figure, performance would be expected to deteriorate when more attributes are monitored.) In a control experiment, observers were asked to monitor one attribute from each object. Here, performance was far worse, showing the difficulty of tracking both objects at once.

In short, attention seems to select all the attributes of one object — even those not immediately relevant to the task in hand — as 'figure'. But it does not seem able to select one attribute of one object, or a mixture of attributes from both objects. This means that attention selects not individual attributes or locations, but rather visual objects as a whole (that is, a set of locations and attributes linked by Gestalt rules). This result is all the more striking because — unlike in some earlier studies — the deck was not stacked in favour of whole-object selection. It would clearly have been advantageous to select a single attribute or a mixture of attributes from each object had it been possible.

How might objects be selected in the visual cortex? And does the apparent restriction to selecting objects — rather than arbitrary sets of attributes — reflect the limitations of neuronal hardware? A conceptually simple scenario is that attention enhances the representation of one object's attributes while attenuating those of the other. But how can attention single out the attributes of just one object? Attributes of the other are encoded in closely overlapping neuronal popula-

tions, and the anatomy of the visual cortex seems to support only relatively coarse and unspecific attentional feedback.

Perhaps one (task-relevant) attribute is selected at first, with feedback spreading to other (non-task-relevant) attributes of the same object through the neural equivalent of Gestalt rules. Alternatively, it is possible that the initial selection is based on the location.

Global change

That sinking feeling

Jorge Sarmiento

The land and sea soak up much of the carbon dioxide emitted into the atmosphere. But one set of simulations suggests that global warming could greatly impair this ability.

Burning of fossil fuels and changes in land use — mainly deforestation — are resulting in more CO₂ in the atmosphere and, it seems, global warming. Much of that extra CO₂ is absorbed in 'sinks' on land and in the oceans. But what effect will future warming have on these sinks? In their paper on page 184 of this issue¹, Cox *et al.* find that in the long run they absorb carbon much less effectively. According to the authors' calculations, the result is 2.5 °C greater global warming over land by the year 2100 than the 5.5 °C predicted if the climate-carbon-cycle connection is not taken into account.

At the moment, the annual increase of CO₂ in the atmosphere is less than half of the estimated emissions². The rest is absorbed by the terrestrial and ocean sinks for carbon. So climate projections have to consider not only future emissions but how those sinks will react^{3,4}. It is no easy matter to couple models of climate change and the carbon cycle, but this is what Cox *et al.* have done.

In their first simulation, they projected how much carbon would be taken up by the

land biosphere and ocean if climate remains constant, as in previous studies. They predefined emissions of CO₂ at the 'business-as-usual' (IS92a) emission scenario⁵. This model predicts that the land biosphere will take up 450 Pg of carbon over the coming century, and the ocean 300 Pg, a grand total of 750 Pg (P is peta, 10¹⁵) (Fig. 1). At an average of 7.5 Pg C yr⁻¹, this is about 50% more per year than the estimated present uptake. The primary mechanism for the land uptake is increased photosynthesis resulting from the increase in atmospheric CO₂ (CO₂ fertilization). In the ocean, it is carbon dissolution of the excess atmospheric CO₂ in the surface waters and transport to depth.

Cox *et al.* then carried out a global warming simulation with atmospheric CO₂ predefined at the IS92a concentrations predicted by a model without climate warming⁵. The reason for doing this was to provide a baseline for how much warming their model would project if there were no feedbacks between climate and the carbon cycle. The projected warming is 5.5 °C over land. A simulation of

It remains to be seen exactly how attention can distinguish between objects represented by populations of neurons that are so intimately entwined. But at the very least, the striking capabilities of visual attention revealed by Blaser *et al.*¹ give us new reasons to think hard about how objects are represented in the visual cortex.

Jochen Braun is at the Institute of Neuroscience,

University of Plymouth, Drake Circus, Plymouth

PL4 8AA, UK.

e-mail: achim@soc.plymouth.ac.uk

1. Blaser, E., Pylyshyn, Z. W. & Holcombe, A. O. *Nature* **408**, 196–199 (2000).

2. Roelfsema, P. R. *et al.* *Nature* **395**, 376–381 (1998).

3. O'Craven, E. M. *et al.* *Nature* **401**, 584–587 (1999).

4. Kapadia, M. K. *et al.* *J. Neurophysiol.* **84**, 2048–2062 (2000).

5. Olshausen, B. A. *et al.* *J. Comp. Neurosci.* **2**, 45–62 (1995).

6. Deneve, S. *et al.* *Nature Neurosci.* **2**, 740–745 (1999).

7. Yeshurun, Y. & Carrasco, M. *Nature* **396**, 72–75 (1998).

8. Lee, D. K. *et al.* *Nature Neurosci.* **2**, 375–381 (1999).

9. Pylyshyn, Z. W. & Storm, R. W. *Spatial Vis.* **3**, 179–197 (1988).

10. Cavanagh, P. *Science* **11**, 1563–1565 (1992).

Model	Mean surface warming in 1860–2100 Global (land)	CO ₂ in 2100 (p.p.m.)	Carbon uptake in 2000–2100 (Pg C)	
			Land	Ocean
1. Predefined CO ₂ emissions without global warming	0.0 °C (0.0 °C)	700	450	300
2. Predefined CO ₂ concentration with global warming	4.0 °C (5.5 °C)	713	-60	250
3. Predefined CO ₂ emissions with global warming	5.5 °C (8.0 °C)	980	-170	400

Figure 1 Global warming and the results of carbon-uptake simulations. Uptake is the amount of carbon absorbed from the atmosphere by 'sinks' on land and in the oceans. Previous studies have typically estimated the concentrations of atmospheric CO₂ for a given emission scenario with models that do not include the effects of global warming on the carbon sinks. The first simulation indicates that such a model based on the IS92a 'business as usual' scenario⁵ puts 750 Pg C into the land and ocean. The second shows that the atmospheric CO₂ projected by such a model will lead to a warming of 5.5 °C over land. But a cross-check on the carbon budget reveals that only 190 Pg C is taken up, mainly because of the effect of warming on the land sink. The third simulation reveals that if the 'missing' 560 Pg C is put back into the model, most of it ends up in the atmosphere. In consequence, warming over land increases to 8.0 °C. (Adapted from ref. 1 with further results from P. M. Cox.)

how much CO₂ the land biosphere and ocean take up with this climate change provides a check on the consistency of the approach of separately simulating climate and the carbon cycle. The result is dramatic. The land biosphere is projected to be a source to the atmosphere of 60 Pg C over the next century, and the uptake by the ocean drops to only 250 Pg. The combined uptake is only 190 Pg C instead of the 750 Pg C predicted by the first study.

Cox and colleagues' final simulation coupled the climate and carbon-cycle models to find out where the carbon goes. The answer is the atmosphere. This simulation projects an atmospheric CO₂ concentration of 980 p.p.m. in 2100, rather than the 700 p.p.m. that comes from the model without climate change. The land biosphere releases 170 Pg of carbon in this model, whereas the ocean takes up 400 Pg of carbon for a total carbon uptake of 230 Pg, only slightly larger than that in the second model. The impact of this increased atmospheric CO₂ on climate is large: warming of 8.0 °C over land, 2.5 °C greater than the climate model with predefined CO₂.

How good are these estimates? The simulations of Cox *et al.* that include climate feedbacks give a larger uptake of carbon by the ocean than those without feedbacks. But this is primarily because the rate of increase of CO₂ in the atmosphere dominates the ocean carbon sink⁶. A comparison of the oceanic uptake between the first and second simulations shows a reduction of more than 20% in the ocean carbon uptake resulting primarily from global warming.

Other models allow further comparisons. A set of simulations with predefined atmospheric CO₂ using the IS92a scenario in four different climate models^{7–10} gave a range in oceanic uptake of 240–470 Pg during the next 100 years (C. Le Quéré, personal communication). The large range of estimates is primarily due to differences in the transport mechanisms that remove carbon from the surface

ocean and carry it to the abyss. But how the ocean circulation and biosphere will respond to global warming is poorly understood. So the finding that these two effects tend to cancel^{3,8,10} must be treated as provisional.

What about the land sink? Cox and colleagues' results stem partly from a dramatic collapse of the Amazonian rainforest because of increased dryness, and partly from a global increase in respiration of organic matter in warmer soils. Here, however, there are especially large uncertainties. For example, a comparative study of the United States¹¹, using various sets of models, estimated that a doubling of CO₂ would change carbon content by anywhere between +32.3% and -39.4%. The lower limit stems from a projected decrease in forested area and increase in water stress; the higher limit from a projected increase in forested area and more vigorous cycling of nitrogen in soils due to warming.

Another group¹² has compared the next generation of land biosphere simulations, in which models of vegetation extent are coupled to models of biogeochemistry. The range of carbon uptake is comparable to the results from the ocean models (240–500 Pg of carbon over the coming century). But only one climate model was used. Moreover, these simulations^{11,12} do not include the direct effects of land vegetation on climate, which tend to increase warming.

But perhaps the greatest source of uncertainty in estimates of the land sink lie in assumptions about CO₂ fertilization and the temperature sensitivity of soil respiration. These responses determine Cox and colleagues' conclusions. But we cannot yet be confident that they will be the dominant mechanisms over the coming century.

For instance, a study¹³ of forests in five US states failed to find substantial growth enhancement from CO₂ fertilization, implying that other processes are counteracting the fertilization effect. Without CO₂ fertilization,

Daedalus

David Jones

David Jones, author of the Daedalus column, is indisposed.

the projected land carbon sink would disappear and soil respiration would dominate. Forest regrowth will help to compensate for this effect. The regrowth sink will decrease with time, however, whereas that due to CO₂ fertilization in Cox and colleagues' model remains strong throughout the century. On the other hand, the temperature sensitivity of soil respiration used in these models, which is the main route for release of carbon, may not hold over the longer timescales of global warming projections¹⁴.

A final illustration of how complex things are comes from another paper in this issue (page 187)¹⁵. Betts suggests that increasing the terrestrial uptake of carbon by planting forests may actually increase global warming in some high-latitude regions. The reason is that darkening of the Earth's surface by trees leads to greater absorption of sunlight.

The bottom line, however, is this. Cox and colleagues' estimate that, by 2100, global warming will result in an extra 600 Pg C in the atmosphere because of a reduction in carbon sinks, an average of 6 Pg C per year. What could this mean in economic terms, given the drive to reduce carbon emissions to the atmosphere? The cost of capturing and storing CO₂ from power plants has been estimated to be about \$200 per ton of carbon emissions avoided¹⁶. Setting aside technological progress that might well reduce this cost, and the possibility of reducing CO₂ emissions, the cost of compensating for a 6 Pg C annual loss in carbon sinks would be around \$1.2 trillion. This calculation, simplistic though it may be, shows what is at stake.

Jorge Sarmiento is in the Program in Atmospheric and Oceanic Sciences, Princeton University, PO Box CN710, Sayre Hall, Princeton, New Jersey 08544-0710, USA.

e-mail: js@princeton.edu

- Cox, P. M., Betts, R. A., Jones, C. D., Spall, S. A. & Totterdell, I. J. *Nature* **408**, 184–187 (2000).
- Houghton, J. T. *et al.* (eds) *Climate Change 1995: The Science of Climate Change* (Cambridge Univ. Press, 1996).
- Sarmiento, J. L. *et al.* *Nature* **393**, 245–249 (1998).
- Cao, M. & Woodward, F. I. *Nature* **393**, 249–252 (1998).
- Houghton, J. T., Callander, B. A. & Varney, S. K. *Climate Change 1992: The Supplementary Report to the IPCC Scientific Assessment* (Cambridge Univ. Press, 1992).
- Sarmiento, J. L. *et al.* *Glob. Biogeochem. Cycles* **9**, 121–138 (1995).
- Bopp, L. *et al.* *Glob. Biogeochem. Cycles* (in the press).
- Joos, F. *et al.* *Science* **284**, 464–467 (1999).
- Maier-Reimer, E. *et al.* *Clim. Dynam.* **12**, 711–721 (1996).
- Matear, R. J. & Hirst, A. C. *Tellus* **51B**, 722–733 (1999).
- VEMAP Members *Glob. Biogeochem. Cycles* **9**, 407–437 (1995).
- Cramer, W. *et al.* *Glob. Change Biol.* (in the press).
- Caspersen, J. *et al.* *Science* (in the press).
- Giardina, C. P. & Ryan, M. G. *Nature* **404**, 858–861 (2000).
- Betts, R. A. *Nature* **408**, 187–190 (2000).
- Ishitani, H. *et al.* in *Climate Change 1995: The Science of Climate Change* (eds Houghton, J. T. *et al.*) 587–647 (Cambridge Univ. Press, 1996).

Invasive alga reaches California

The alga has been identified that threatens to smother Californian coastal ecosystems.

The recent discovery of the marine green alga *Caulerpa taxifolia* on the Californian coast^{1,2} has raised public concern about the potential danger of a new invasion similar to the one endured by the Mediterranean Sea over the past decade. A small colony of *C. taxifolia* introduced into the Mediterranean in 1984^{3,4} from a public aquarium⁵ has spread to more than 6,000 hectares today, outcompeting native species and seriously reducing diversity in areas of the northwestern Mediterranean⁶. This invasive strain of *C. taxifolia* differs from tropical populations in that it is much larger, grows more vigorously, does not rely on sexual reproduction, and is resistant to low temperatures^{4,6,7}. Here we evaluate the risk of invasion by Californian *C. taxifolia* by comparing it genetically with the Mediterranean and aquarium strain, as well as with native tropical populations. Our results show that the Californian alga is the same as the invasive Mediterranean strain, calling for its rapid eradication to prevent a new invasion.

Colonies of *C. taxifolia*, morphologically similar to the Mediterranean strain, have been reported from Carlsbad and Huntington Harbour, California. *C. taxifolia* covers 3,500 square metres at Carlsbad and is dispersed over 20,000 square metres at Huntington Harbour. We extracted DNA from algal samples collected in these areas in independent analyses in two laboratories (Geneva, Switzerland, and Fresno, California). The internal transcribed spacer of ribosomal DNA (ITS rDNA) was amplified by using the polymerase chain reaction and then cloned⁵, and the sequence data were used to examine intraspecific variability within *C. taxifolia*. Five clones were sequenced for each isolate in Geneva; two specimens analysed in Fresno were sequenced directly. We compared Californian alga sequences with 148 ITS rDNA sequences obtained from 32 *C. taxifolia* specimens collected from all over the world.

Eleven out of twelve Californian sequences were found to be identical to all aquarium and most Mediterranean sequences (Fig. 1a). The single divergent sequence (Carl4), which differs by two nucleotide substitutions, branches with other Mediterranean sequences. This group is clearly separated from all sequences obtained from other isolates of *C. taxifolia*, including tropical strains of this species collected in the Red Sea, Caribbean Sea and the Indo-Pacific (Fig. 1b). Phylogenetic analysis of all *C. taxifolia* ITS rDNA sequences reveals a relatively robust clade (80% bootstrap value) grouping Californian, Mediterranean, aquarium and some

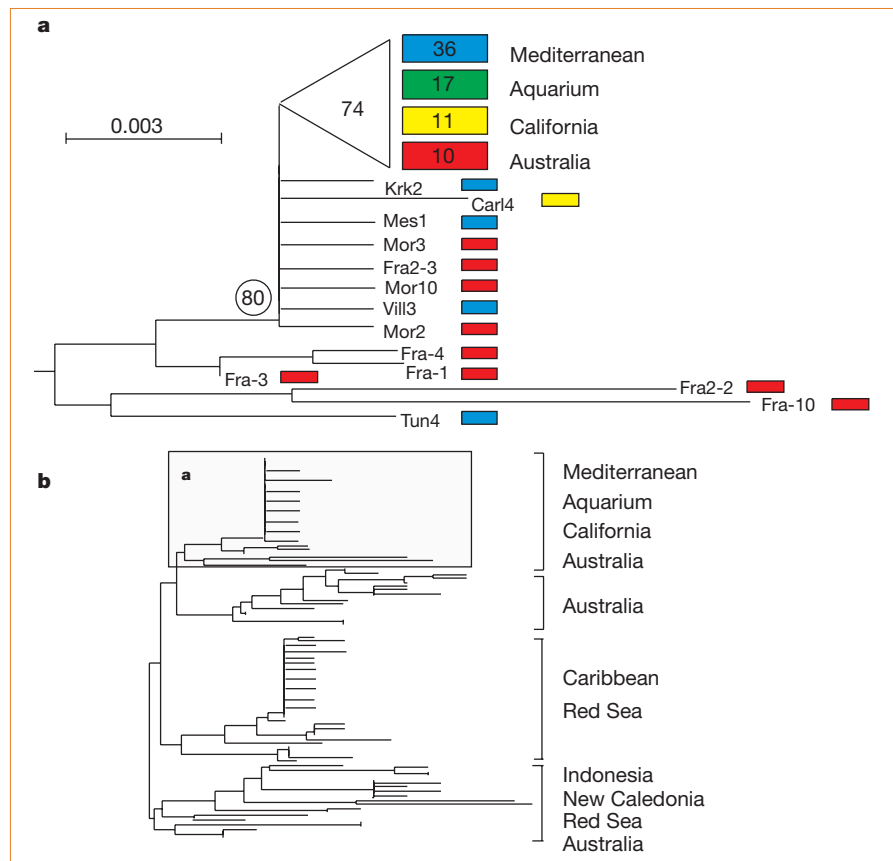


Figure 1 Phylogenetic tree of *C. taxifolia* inferred from 160 complete ITS (internal transcribed spacer) ribosomal DNA sequences by using the neighbour-joining method applied to pairwise sequence distances calculated by the Kimura two-parameter method. **a**, The clade containing all Californian sequences; the origin of different sequences is indicated by coloured rectangles containing the number of identical sequences; number 80 at the base of the clade corresponds to the bootstrap value based on 1,000 replicates. Abbreviations are: Krk (Krk Island, Croatia), Carl (Carlsbad, California), Mes (Messina, Italy), Mor (Moreton Bay, Australia), Fra (Fraser Island, Australia), Vill (Villefranche, France), Tun (Sousse, Tunisia). **b**, Total tree, presenting relationships between algae of different geographical origins. Sampling localities (number of analysed sequences in parentheses) in the Mediterranean Sea: Majorca (3), Saint-Cyprien (3), Le Brusc (3), Toulon (3), Port-Cros (3), Le Lavandou (3), Villefranche (5), Messina (3), Elba (3), Krk (3) and Hvar (3) Islands, Croatia; Sousse, Tunisia (5); aquarium samples: Nancy (3), Stuttgart (5), Geneva (3), Oahu (3), Enoshima (3); sampling localities off the Californian coast: Carlsbad (6) and Huntington Harbour (6); from Australian waters: Townsville⁵ (1), Fraser Island (13), Moreton Bay (18), Port Hacking (5), Lake Conjola (9); in the Caribbean: Puerto-Rico (5), Martinique (7), Guadeloupe (3); from the Red Sea: Safaga, Egypt (21); from Indonesian waters: Djakarta⁹ (1); and from New Caledonia: Noumea (8). Ethanol-fixed specimens of Californian *C. taxifolia* have been deposited in the herbarium of the Botanical Garden in Geneva and in the Jepson Herbarium at the University of California, Berkeley. EMBL accession numbers for the sequences analysed are AJ228960–AJ228999 and AJ299742–AJ299811.

ian, Mediterranean, aquarium and some Australian sequences. Thus, the Californian alga belongs to the aquarium–Mediterranean strain.

A particular feature of all aquarium, Mediterranean and Californian (AMC) specimens we examined is the low degree of intra-individual polymorphism of their rRNA genes: 64 sequences from AMC isolates were identical, whereas five were slightly divergent (from 0.4 to 1.1%). Such homogeneity contrasts with the high polymorphism of ITS copies in non-AMC isolates, characterized by intra-isolate sequence divergence ranging from 0.8 to 3.6%.

It is notable that the eastern Australian polymorphic isolates possess ITS copies that are identical to the AMC type (10 out of 46 Australian sequences). This may point to Australia as being the possible native land of the invasive alga, but this needs to be confirmed by analysis of other genetic markers to ascertain the similarity of the strains.

O. Jousson*, J. Pawlowski*, L. Zaninetti*, F. W. Zechman†, F. Dini‡, G. Di Guiseppe‡, R. Woodfield§, A. Millar||, A. Meinesz¶

* Université de Genève, Département de Zoologie et Biologie Animale, 1224 Chêne-Bougeries, Switzerland
e-mail: jousson2@sc2a.unige.ch

†Department of Biology, California State University, Fresno, California 93740, USA
‡Università di Pisa, Dipartimento di Etologia, Ecologia ed Evoluzione, 56126 Pisa, Italy
§Merkel and Associates, San Diego, California 92123, USA
||Royal Botanic Gardens, Sydney, New South Wales 2000, Australia
¶Université de Nice-Sophia Antipolis, Laboratoire Environnement Marin Littoral, 06108 Nice, France

Imaging

Phase radiography with neutrons

The interaction of neutrons with matter enables neutron radiography¹ to complement X-ray radiography in analysing materials. Here we describe a simple quantitative method that provides a new contrast mechanism for neutron radiography and allows samples to be imaged at low radiation doses. Large phase shifts can be measured accurately from detailed structures not amenable to conventional techniques.

The complexity of interferometry has limited studies on neutron phase imaging. Our non-interferometric phase-measurement technique is appropriate for non-uniform radiation and returns a unique, non-wrapped, smooth, continuous quantitative phase map; it implements a theoretical study² and has been used for other types of radiation, including visible light³, electrons⁴, soft⁵ and hard⁶ X-rays.

We base our analysis on the paraxial transport of intensity equation^{7,8}

$$\frac{2\pi}{\lambda} \frac{\partial I(\mathbf{r}_\perp)}{\partial z} = -\nabla_\perp \times [I(\mathbf{r}_\perp) \nabla_\perp \phi(\mathbf{r}_\perp)] \quad (1)$$

where the wave is described by $\sqrt{I(\mathbf{r}_\perp)} \exp(i\phi(\mathbf{r}_\perp))$, with intensity $I(\mathbf{r}_\perp)$ and phase $\phi(\mathbf{r}_\perp)$; $\nabla_\perp$ and $\mathbf{r}_\perp$ are respectively the gradient operator and position vector in the plane perpendicular to the longitudinal optic axis z , λ is the wavelength, and $\partial I(\mathbf{r}_\perp)/\partial z$ is the longitudinal intensity derivative.

We used the beam port of the Neutron Interferometry facility at NIST, operating at a selectable monochromatic wavelength of 4.43 Å. The divergence of the neutron beam was limited by a mask to around 6 milliradians and neutrons illuminated a sample at about 1.8 m from the beam-guide exit. The two-dimensional detector⁹ was 30% efficient and combined an in-beam ⁶Li-ZnS scintillator, light intensifier and mirror. The mirror reflected the light out of the neutron beam and into an optical charge-coupled device camera. Electronic centroiding of images optimized the spatial resolution to 60 μm.

The intensity over a surface is consistent

1. Kaiser, J. *Science* **289**, 222–223 (2000).
2. Dalton, R. *Nature* **406**, 447 (2000).
3. Meinesz, A. & Boudouresque, C. F. C. *R. Acad. Sci. Paris Life Sci.* **319**, 603–613 (1996).
4. Meinesz, A. & Hesse, B. *Oceanol. Acta* **14**, 415–426 (1991).
5. Jousson, O., Pawlowski, J., Zaninetti, L., Meinesz, A. & Boudouresque, C. F. *Mar. Ecol. Prog. Ser.* **172**, 275–280 (1998).
6. Meinesz, A. *et al.* *Bot. Mar.* **38**, 499–508 (1995).
7. Komatsu, T., Meinesz, A. & Buckles, D. *Mar. Ecol. Prog. Ser.* **146**, 145–153 (1997).
8. Olsen, J. L., Valero, M., Meusnier, I., Boele-Bos, S. & Stam, W. T. J. *Phycol.* **34**, 850–856 (1998).
9. Pighini, M. & Dini, F. *4th International Workshop on Caulerpa taxifolia* (Lerici, Italy, 1999).

with any phase distribution. Equation (1) indicates, however, that the phase does influence the propagation of the intensity. It is therefore possible to visualize phase in a certain plane by allowing a wave to evolve

away from that plane¹⁰ — an example of this is the familiar heat shimmer over a hot road. This effect was used to visualize the neutron phase shift for a North American yellow-jacket wasp (genus *Vespula*) (Fig. 1). A conventional neutron radiograph (Fig. 1a) is compared with a phase-contrast image obtained 1.8 m downstream from the sample using neutrons emanating from a 0.4-mm-diameter pinhole placed at the exit of the beam guide (Fig. 1b). The pinhole sets the limit to the spatial resolution. The improved sensitivity of the image arises from shifts in the neutron phase introduced by the sample.

To retrieve phase quantitatively using equation (1), we require the intensity $I(\mathbf{r}_\perp)$ and intensity derivative $\partial I(\mathbf{r}_\perp)/\partial z$ in the plane of interest. To obtain this, images were

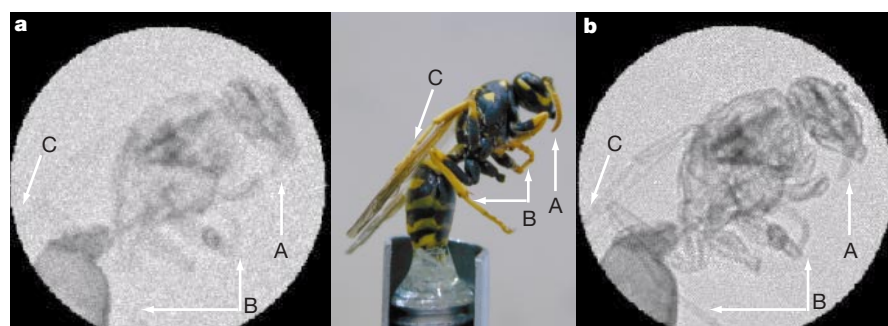


Figure 1 Neutron phase-contrast imaging of a yellow-jacket wasp. **a**, Conventional neutron radiograph. **b**, Phase-contrast image showing greater clarity of finer features such as the antenna (A), legs (B) and projection through the leg and wing (C).

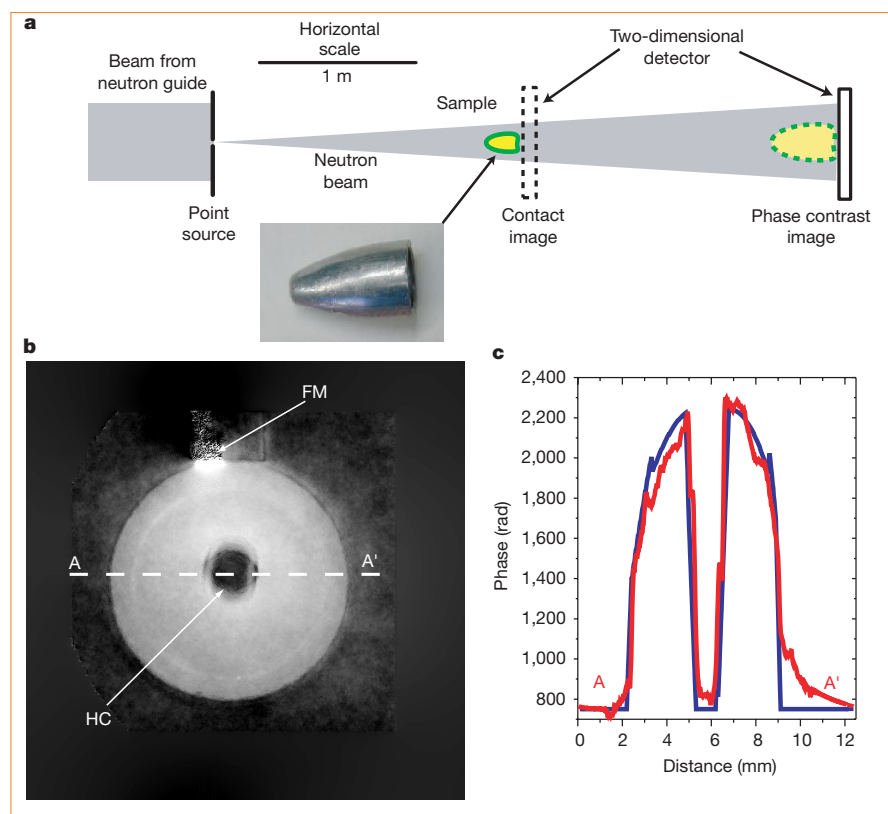


Figure 2 Quantitative neutron phase imaging. **a**, Experimental scheme, with the lead sinker sample shown in the inset. **b**, Quantitative phase map of the Pb sinker showing the hollow core (HC) and fiducial mark artefact (FM). **c**, Quantitative phase profile AA' through the sinker (red) and calculated profile (blue) based on the known shape, scattering length and orientation of the sinker.

taken in two planes: the first was a contact image and the second was a phase-contrast image with the detector 1.8 m from the object (Fig. 2a). Using the contact image and the phase-contrast image, we approximate the intensity by their average, and the intensity derivative by their difference divided by their separation. We used this approach to image a sinker made of lead aligned longitudinally with the beam (Fig. 2a, inset). The measured phase from equation (1) is shown in Fig. 2b. The phase deformity is an artefact of a gadolinium fiducial mark; the hollow core can be clearly seen in the phase image.

The phase image obtained is, to a good approximation, described by a convolution of the perfect image with the intensity distribution of the effective source. In our analysis, we used a very conservative estimate of the effective width of the neutron source to modify the phase-recovery algorithm. A profile of the recovered phase is plotted in Fig. 2c, together with the predicted phase profile determined from the sample geometry, scattering length and orientation, and gives good quantitative agreement between the two. Note that the 1,400-rad phase excursion of the sample occurs in a few hundred micrometres (less than ten pixels). An interferometric experiment would require submicrometre resolution to measure such a rapid phase excursion accurately.

This simple and general technique provides independent quantitative phase and absorption images of the sample. These images constitute two-dimensional projections of the real and imaginary parts of the neutron refractive-index distribution through a three-dimensional object. A number of these projections could be used to generate a quantitative tomographic reconstruction.

B. E. Allman*, P. J. McMahon*, K. A. Nugent*, D. Paganin*, D. L. Jacobson†, M. Arif†, S. A. Werner‡

*School of Physics, University of Melbourne, Victoria 3010, Australia

e-mail: k.nugent@physics.unimelb.edu.au

†Physics Laboratory, National Institute of Standards and Technology (NIST), Gaithersburg, Maryland 20899, USA

‡Department of Physics and Astronomy, University of Missouri-Columbia, Columbia, Missouri 65211, USA

- Schlenker, M. & Baruchel, J. *Physica B* **137**, 309–319 (1986).
- Paganin, D. & Nugent, K. A. *Phys. Rev. Lett.* **80**, 2586–2589 (1998).
- Barty, A., Nugent, K. A., Paganin, D. & Roberts, A. *Opt. Lett.* **23**, 817–819 (1998).
- Bajt, S. *et al. Ultramicroscopy* **83**, 67–73 (2000).
- Allman, B. E. *et al. J. Opt. Soc. Am. A* **17**, 1732–1743 (2000).
- Nugent, K. A., Gureyev, T. E., Cookson, D., Paganin, D. & Barnea, Z. *Phys. Rev. Lett.* **77**, 2961–2964 (1996).
- Teague, M. R. *J. Opt. Soc. Am.* **73**, 1434–1441 (1983).
- Gureyev, T. E., Roberts, A. & Nugent, K. A. *J. Opt. Soc. Am. A* **12**, 1942–1946 (1995).
- Dietze, M., Felber, J., Raum, K. & Rausch, C. *Nucl. Inst. Meth. A* **377**, 320–324 (1996).
- Zernike, F. *Physica* **9**, 686–693 (1942).

Scaling

Rivers, blood and transportation networks

The search for a theory to explain why the metabolic rate of mammals is proportional to the 3/4-power of body mass (Kleiber's law) has recently focused on the nutrient distribution network formed by arteries and capillaries. Banavar *et al.*¹ argue that the law follows from the intrinsic properties of an outward-directed network. But careful analysis of their arguments reveals two implicit assumptions that may not be generally correct. Unless these assumptions are valid for mammalian circulation, these arguments cannot satisfactorily explain Kleiber's empirical relationship.

In the analysis by Banavar *et al.*¹, there is a site for nutrient uptake at each network-branching point, and the distance L_X along a path from the origin O (the heart) to a site X is defined as the number of uptake sites on the path. The rate of uptake of nutrients at a site is a constant F_X (assumed to be species- and size-invariant). A network segment that goes from a site X to an adjacent site Y is termed the link XY , and the rate at which nutrient enters the link is termed the current and is denoted by I_{XY} . In an outward-directed network (ODN), the direction of flow is away from O on each link.

The authors' fundamental result is that, in an ODN, $C = B \langle L_X \rangle$, where C , the total current in the network, is the sum of currents on all links, B is the metabolic rate (sum of uptake rates), and $\langle L_X \rangle$ is the average distance to sites. The number of uptake sites is expressed as L^3 , and the metabolic rate B is therefore $F_X L^3$. Banavar *et al.* prove that, if $\langle L_X \rangle$ is proportional to L , if C is proportional to blood volume V_b , and if V_b is proportional to body size M , then B is proportional to $M^{3/4}$, which is Kleiber's law.

Although the assumption that $V_b \propto M$ is a reasonable approximation, the assumption that $C \propto V_b$ is not, in general, true. Consider a network formed by $(l-1)(m-1)(n-1)$ unit cubes stacked to form a rectangle $(l-1)$ units deep by $(m-1)$ units wide by $(n-1)$ units high. Each vertex of a cube is an uptake site and each edge is a network link with constant cross-sectional area. Current is supplied to the bottom rear left corner of the network and flows forward, rightward and upward. It follows from induction on n that the sum of currents in all vertical links is $lmn(n+1)F_X/2$. Adding the corresponding expressions for currents flowing rightward and forward gives $lmnF_X(l+m+n+3)/2$, the total current C . The blood in the network is proportional to the number of links, $3lmn - lm - ln - mn$. Thus, in this model, in which the density of uptake sites is invariant, C is nearly proportional to volume multiplied by average linear dimension, metabo-

lism is proportional to volume, and V_b is approximately proportional to volume.

The relation between V_b and C depends on parameters that do not appear in Banavar *et al.*'s report¹. The volume of blood in link XY is equal to the cross-sectional area of the link (A_{XY}) multiplied by the length of the link (s_{XY}), whereas current I_{XY} is A_{XY} multiplied by flow velocity (f_{XY}). The volume of blood in a link is therefore $I_{XY}s_{XY}/f_{XY}$. The assumption that the sum of currents equals (or scales as) total blood volume implies that the average value of s_{XY} increases as size increases. In arteries, the flow velocity may increase as size increases, but flow rates in capillaries are severely constrained by their relatively constant cross-sectional area and pressure. If most uptake sites are located in capillary networks, the currents in capillaries comprise most of the sum C , but the blood in these vessels may comprise a minority of blood volume.

The assumption that $\langle L_X \rangle \propto L$ is also not true in general in an ODN. For example, consider a network that starts with a single link and bifurcates at each branch point until a terminal uptake site is reached at distance k . The number of uptake sites, L^3 , is $2^k - 1$ and $\langle L_X \rangle$ is $[(k-1)2^k - 1]/(2^k - 1)$. Therefore $\langle L_X \rangle$ is approximately proportional to the logarithm of L in this type of ODN.

Page R. Painter

Office of Environmental Health Hazard Assessment, California Environmental Protection Agency, 301 Capitol Mall, Sacramento, California 95814-4327, USA
e-mail: ppainter@oehha.ca.gov

The long-standing problem of explaining metabolic scaling² in animals, whereby whole-animal metabolic rate B is observed to increase as a function of body mass M approximately as $M^{3/4}$, has been recently revisited by Banavar *et al.*¹ (see also ref. 3, in which allometric scaling rules are derived from fractal geometry). These authors¹ derive and generalize to non-biological systems, including river networks, a three-quarter-power 'allometric' scaling rule, which arises, in their treatment, from an assumption of the efficiency of the resource distribution network. Here I present a simple derivation of 3/4-power scaling based on the geometric requirements of inventorying resources before metabolization, which does not support the notion of allometric scaling suggested by Banavar *et al.*¹ for rivers, at least not when applied to the problem of fluvial sediment transport. Although some distributary systems 'metabolize' according to the 3/4-power rule, this rule is not golden — each system needs to be investigated on its own merits.

As a resource distribution and processing (metabolizing) system increases in size, there is a geometric necessity for an incompressible and conserved resource for

resource inventory in the transportation network also to increase as it awaits metabolization. Equating the rate of flow through the inventory volume to the rate of flow through the metabolizing volume provides a relation determining the scaling behaviour of the metabolization rate. Thus, if a three-dimensional metabolizing system contains two volumes, one, V_M , within which metabolization of resources occurs, and one, V_I , within which the resource is inventoried, and if these volumes are sequentially connected and the resource is incompressible and conserved, then $V_M/\tau = V_I/T$, where τ is the residence time of resource in the metabolizing volume, and T is the residence time of the resource in inventory. The metabolic rate is $B \propto V_M/\tau$.

If metabolization occurs in small volumes, V_{met} , of fixed size, then $V_M \propto V_{met}(L/l)^3$, where l is the average spacing between metabolizing volumes. The inventoried resource for each metabolizing volume V_{met} is distributed in a queue that extends across the length of the system L , with L/l units of inventory volume V_I earmarked for each volume V_{met} . If V_I is independent of L (which would be expected in systems where the rate-limiting process was metabolization rather than transport), then T/τ is proportional to L/l (that is, the ratio of average velocities U/u in the two volumes V_I and V_M is independent of L). The total inventory volume then scales as $V_I \propto (L/l)^4$ (that is, $l \propto L^{1/4}$), so that $L/l \propto V_I^{1/4}$, and metabolic rate is given by $B \propto V_I^{3/4}/\tau$.

Three-quarter-power scaling of total metabolic rate as a function of system size follows under the additional assumptions¹ that V_I ('blood' volume) is proportional to system ('whole animal') volume (or mass), and that τ is independent of system size. It is worth noting that, in the geometric argument presented here, the distribution of metabolic units, V_{met} , is taken to be uniform for purposes of counting, but the actual spacing can be highly non-uniform without affecting the overall scaling argument.

If allometric scaling of the kind found for animals has almost universal applicability to resource distribution systems, as suggested by Banavar *et al.*¹, it ought to apply to rivers, an application they consider. Consider the particular case of sediment transport in rivers, in which the metabolic rate of a drainage basin of area A is the rate at which sediment is delivered by the trunk stream to the lower end of the drainage basin. An analogue of blood volume V_I is the total sediment volume in transport, $V_{sed}(L/l)^2(T/\tau)$, where V_{sed} is the volume of sediment delivered from hillslopes to stream channels from a small hillslope area l^2 during time τ , and T is the residence time of such sediment in the channel system of drainage basin A with linear dimension L .

However, in fluvial systems there is no

apparent basis for assuming that the velocity ratio U/u is independent of A . It is also difficult to identify a fluvial analogue of the constraint¹ that blood volume is proportional to animal volume. One difference between rivers and organisms in this regard is that, in rivers, sediment can be easily stored laterally and vertically in flood plains. Flood plains, lying outside the channels themselves, occupy a third volume representing a kind of long-term storage. Such 'loss' of sediment between hillslope and river mouth can be a significant effect⁴ in fluvial sediment budgets. The relation equating rates of flow in inventory and metabolizing volumes is valid only if there is no third volume for storing resources. Models of the kind proposed by Banavar *et al.*¹ for biological organisms do not appear to be sufficient for deriving scaling relations for river (sediment) metabolism; such models potentially apply only to distribution of an incompressible resource in a system with 'two-volume' geometry.

Peter K. Haff

Division of Earth and Ocean Sciences, Nicholas School of the Environment, Duke University, Durham, North Carolina 27708-0230, USA
e-mail: haff@duke.edu

Banavar et al. reply — The idea behind our theorem¹ is simple. It can be illustrated by using airline travel as an example. Consider a stream of people (blood) leaving London (heart) at a steady rate and fanning out to all parts of the world (body). The number of people leaving London each day and arriving elsewhere at their final destinations (metabolic rate) is denoted by B . Assuming that the people travel along a locally connected network and that the transit time for each local hop is the same (say, 1 day), the number of people in transit at any given time (blood volume) is proportional to B , but with a proportionality constant that is given by the mean number of hops from all the destination cities to London.

This additional factor arises because if, for example, there are P people who arrive in, say, Paris (let us assume Paris is two hops from London — London–Brussels–Paris) each day, there are $2P$ passengers in transit at any given time whose final destination is Paris (P of them are en route from London to Brussels and the other P are travelling from Brussels to Paris). In a D -dimensional space, if B scales as L^D , our theorem asserts that because the mean number of hops must itself scale at least as L , the total number of people in transit must scale at least as L^{D+1} .

Haff points out the difference between the scaling properties of water flow and sediment transport in rivers and that river sediments can be stored in flood plains. Although water flow at a given point is proportional to the area of the sub-basin draining into it, the sediment discharge is

not, because the source of sediments is not uniform in space, unlike the rainfall in landscape-forming events⁵. Rather, sediment production is scattered in space and time, and not isochrone with the main transport mechanisms in the network. In fluvial systems, scaling networks are stationary structures derived from the evolutionary dynamics of the topography of landscapes⁶.

Painter's exercise of the parallelepiped, showing that the mean number of hops from the origin, $\langle L_x \rangle$, is proportional to L is in accord with our theorem. Painter assumes that the blood in the network is proportional to the number of links — or, by analogy, that the total number of people in transit is proportional to the number of operating flights — but this assumption is wrong because there are many more people crammed into flights originating from London than in those from cities distant from London. The number of flights is indeed proportional to the number of destinations and the metabolic rate B . The branching network in Painter's last paragraph is a Cayley tree which, for large enough sizes, cannot exist in any finite-dimensional space. The observation that in an N -site Cayley tree the average distance from the origin scales as $\ln N$, which is what would be expected for a D -dimensional lattice in the infinite D limit, agrees with our theorem.

We do not believe that fractal-like networks effectively endow life with an additional fourth dimension³. Allometric scaling comes built in with any system in which the flow is directed and the circulation time is proportional to circulation length, irrespective of size¹. The fact that nature, in spite of her extreme diversity, exhibits allometric scaling to the extent she does in plants, animals and river networks suggests that optimality associated with directedness is quite common.

Jayanth R. Banavar*, **Amos Maritan†**, **Andrea Rinaldo‡**

**Department of Physics and Center for Materials Physics, 104 Davey Laboratory, Pennsylvania State University, University Park, Pennsylvania 16802, USA*

e-mail: banavar@psu.edu

†International School for Advanced Studies, Via Beirut 2-4, 34014 Trieste, and INFN and the Abdus Salam International Center for Theoretical Physics, 34014 Trieste, Italy

‡Dipartimento IMAGE, Università di Padova, via Loredan 20, 35131 Padova, Italy

1. Banavar, J. R., Maritan, A. & Rinaldo, A. *Nature* **399**, 130–132 (1999).
2. Schmidt-Nielsen, K. *Scaling: Why is Animal Size so Important?* (Cambridge Univ. Press, 1984).
3. West, G. B., Brown, J. H. & Enquist, B. J. *Science* **284**, 1677–1679 (1999).
4. Ritter, D. F., Kochel, C. R. & Miller, J. R. *Process Geomorphology* 3rd edn (WCB/McGraw-Hill, Boston, 1995).
5. Rodriguez-Iturbe, I. & Rinaldo, A. *Fractal River Basins: Chance and Self-Organization* (Cambridge Univ. Press, New York, 1997).
6. Banavar, J. R. *et al. Phys. Rev. Lett.* **78**, 4522–4525 (1997).

Chemical and biological trends during lake evolution in recently deglaciated terrain

Daniel R. Engstrom*, Sherilyn C. Fritz†, James E. Almendinger* & Stephen Juggins‡

* St Croix Watershed Research Station, Science Museum of Minnesota, Marine on St Croix, Minnesota 55047, USA

† Department of Geosciences and School of Biological Sciences, University of Nebraska, Lincoln, Nebraska 68588-0340, USA

‡ Department of Geography, University of Newcastle, Newcastle upon Tyne NE1 7RU, UK

As newly formed landscapes evolve, physical and biological changes occur that are collectively known as primary succession. Although succession is a fundamental concept in ecology, it is poorly understood in the context of aquatic environments. The prevailing view is that lakes become more enriched in nutrients as they age, leading to increased biological production. Here we report the opposite pattern of lake development, observed from the water chemistry of lakes that formed at various times within the past 10,000 years during glacial retreat at Glacier Bay, Alaska. The lakes have grown more dilute and acidic with time, accumulated dissolved organic carbon and undergone a transient rise in nitrogen concentration, all as a result of successional changes in surrounding vegetation and soils. Similar trends are evident from fossil diatom stratigraphy of lake sediment cores. These results demonstrate a tight hydrologic coupling between terrestrial and aquatic environments during the colonization of newly deglaciated landscapes, and provide a conceptual basis for mechanisms of primary succession in boreal lake ecosystems.

Eutrophication is often described as an acceleration of the natural ageing process of lakes, based on an early ecological model in which lakes become more productive as they age¹. This model of lake development, which originated with efforts to classify European lakes by trophic status^{2–4}, persists to the present day⁵ despite palaeolimnological evidence that many temperate-region lakes follow the opposite trend and become more dilute and unproductive over time^{6–10}. Although sedimentary records have advanced our understanding of lake development beyond the simple eutrophication model, the patterns and controls of long-term limnological change remain poorly understood, because direct observation is impossible at these timescales and because few regions exist where lakes can be studied immediately following formation. In glaciated landscapes, where most of the world's lakes occur, the period following ice recession saw dramatic changes in terrestrial vegetation, soils and hydrology. Yet we know little of how newly formed glacial lakes respond to these environmental forces, whether there are common pathways of development within a given geographical region, or how underlying differences in geology and climate affect the direction and rate of limnological change. Moreover, it is unclear whether changes to lakes ('lake trajectories') are in fact driven by ecological processes set in motion by glacial retreat or whether they represent the direct forcing of a rapidly changing climate¹¹.

To explore more directly the environmental forces controlling lake development, we studied a suite of recently formed lakes in Glacier Bay National Park, Alaska, where more than 1,000 years of late Holocene ice advance and retreat has created lake-studded glacial forelands, among the most extensive found anywhere in the world¹². We compare limnological conditions among 33 lakes of differing ages, and infer patterns of limnological change by assuming that the sequence represents lakes at different developmental stages. This approach has been widely used to model long-term developmental processes in terrestrial systems, as exemplified by the classic studies of primary succession at Glacier Bay^{13–17}. Because a spatial array of modern sites may not fully describe temporal patterns exhibited by a single site over its history^{18–20}, we also reconstruct limnological trends from sediment cores from individual lakes and

compare these trends with those exhibited by the chronosequence. Lake histories are reconstructed from fossil diatoms by using transfer functions developed from the relationship between modern diatom assemblages and environmental variables in the Glacier Bay lakes.

The Glacier Bay chronosequence

The studied lakes were created by a sequence of glacial advances and retreats dating from this century back to late Wisconsin time (14,000 years ago). All occupy small primary catchments receiving no drainage from other lakes or major streams; most are small (3–16 ha) and moderately deep (maximum depth, 3–18 m). Twenty-one were formed by catastrophic recession of late Holocene ice from low-elevation forelands along the Glacier Bay fjord, and range in age from 10 to 220 years (Fig. 1). Twelve additional lakes are located on glacial forelands along the outer coast (350–2,700 years) and on late Wisconsin surfaces on Pleasant Island (13,000 years). Catchment vegetation grades from a sparse cover of early colonizers (*Epilobium*, *Dryas*, *Salix*) at the youngest sites (<50 years old) to dense shrub thickets of Sitka alder (*Alnus sinuata*), closed forests of Sitka spruce (*Picea sitchensis*) and western hemlock (*Tsuga heterophylla*), and ultimately open peatlands on the oldest surfaces (>1,000 years old). The Glacier Bay region has a maritime climate with small annual temperature variations, frequent cloud cover and heavy precipitation^{21,22}.

Each lake in the chronosequence was sampled for water chemistry a minimum of three times over a three-year period to assess seasonal and interannual variability; these data are supplemented for 17 sites by collections from three previous years. The lakes exhibit a wide range of ionic strengths (ion sums, 0.1–6.2 mequiv. l⁻¹); the more concentrated lakes (ion sums, >1 mequiv. l⁻¹) are dominated by Ca²⁺ and HCO₃⁻ ions, whereas the more dilute lakes have Na⁺, Cl⁻ and organic anions in greater proportion. Concentrations of chlorophyll *a* are below 1.2 parts per billion (p.p.b.; median, 0.3 p.p.b.), total phosphorus (TP) less than 10.5 p.p.b. (median, 5.5 p.p.b.), and total nitrogen (TN) less than 400 p.p.b. (median, 215 p.p.b.), indicating oligotrophic conditions throughout the chronosequence.

Most limnological parameters exhibit clear trends with lake age.

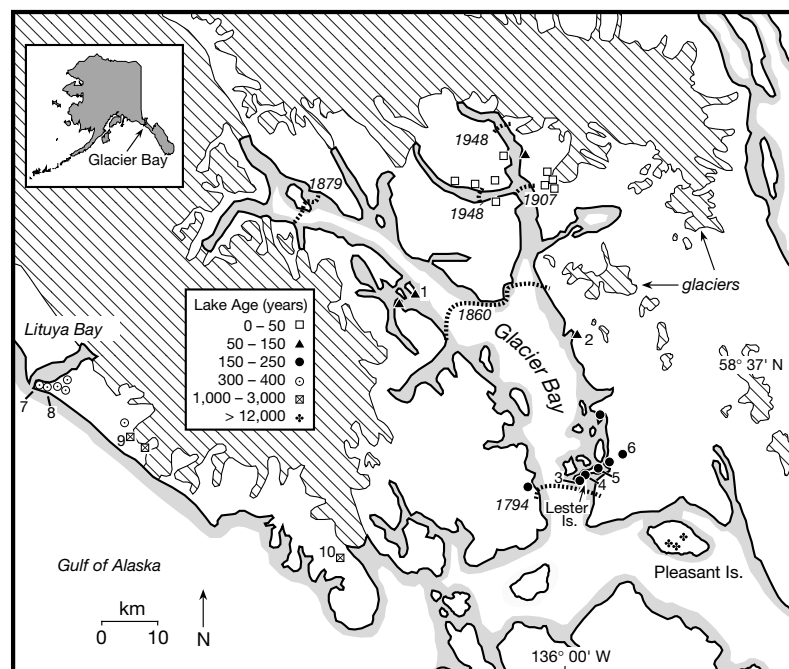


Figure 1 A chronosequence of lakes in the Glacier Bay region of southeastern Alaska. Late Holocene ice margins are marked by dated isochrons; numbered lakes have core trajectories shown in Fig. 3. Lake age along the Glacier Bay fjord is generally contemporaneous with the date of surface deglaciation, which is known from the observations of early explorers and scientists, tree-ring chronologies, and photographic

records¹⁶. The ages of lakes at Lituya Bay are based on ¹⁴C dating of morainal surfaces⁴⁰, while the ages of older sites along the outer coast and on Pleasant Island were established by ¹⁴C dating of basal sediments from cores obtained in this study (Brady, 1230 ± 50 years BP; Dagelet, 1080 ± 60 years BP; LaPerouse, 2690 ± 60 years BP) and from Engstrom *et al.*⁴¹.

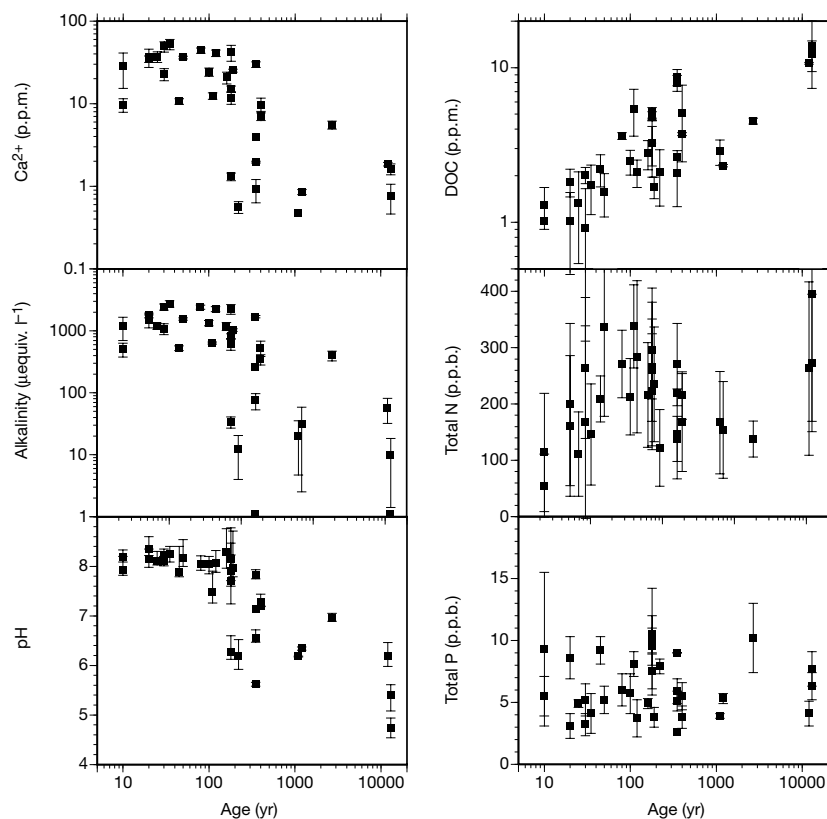


Figure 2 Relationship between lake age and mean values of selected water-chemistry variables in the Glacier Bay chronosequence. Error bars (± 1 s.d.) depict variation among sampling dates. Lakes were sampled at a depth of 1 m with a Kemmerer bottle; pH was measured in the field; alkalinity (Gran titration) and apparent colour were measured within 12 h of sample collection. Samples were filtered (at 0.45 μm) and frozen for laboratory

analysis of DOC (carbon analyser) and major anions (ion chromatography). Unfiltered samples were analysed for TP and TN (colorimetric methods) and major cations (direct current plasma-atomic emission spectrometry) TN samples were acidified. Chlorophyll *a* was measured by high-performance liquid chromatography on samples filtered in the field.

All major cations and all anions except Cl^- show declining concentrations along the chronosequence of sites from youngest to oldest. The general trend is illustrated here (Fig. 2) by the results for Ca^{2+} and alkalinity. While there is substantial scatter in the data for any particular time slice, there are clearly no old lakes with high Ca^{2+} or alkalinity, and conversely no young lakes with low Ca^{2+} or alkalinity. Trends are also evident in pH, dissolved organic carbon (DOC) and TN. Surface-water pH is uniformly high (> 8.0) for lakes less than 200 years old, but shows a declining pattern with lake age among older sites (Fig. 2). This pH trend reflects the greater carbonate buffering of the younger lakes (high alkalinities) and the increasing concentration of organic acids (as illustrated by DOC) in the older lakes. Total N also increases with lake age, but only for the first one to two centuries, declining thereafter except for the three oldest lakes on Pleasant Island. Total-P concentrations, on the other hand, show no clear trend with age since deglaciation.

An analysis of diatoms in surface sediments from the 33 lakes suggests that changes in water chemistry over time drive changes in biological communities. Detrended canonical correspondence analysis (DCCA) of the diatom assemblages and associated water-chemistry data²³ show that the major axis of variation in the diatom data is closely aligned with pH and DOC, while the second axis is most strongly correlated with TN (Fig. 3 inset). Axis 3 (not shown) reflects the influence of sea-spray (Na^+ , Cl^-). Diatom assemblages in lakes of similar age are of similar composition: the youngest lakes (< 50 years) cluster in the lower right of the diagram (high pH, low DOC, low TN), lakes of intermediate age (50–250 years) generally cluster in the upper right (high pH, low DOC, medium to high TN), and the older lakes (> 300 years) cluster on the left side of the first

axis (lower pH, high DOC, medium TN).

Sediment records of lake trajectories

To corroborate the pattern of limnological development suggested by the chronosequence, we analysed fossil diatoms in sediment cores from 10 individual lakes ranging in age from approximately 100 to 1,000 years. These cores were collected from the deepest region of each lake, sectioned in the field at 0.5–1.0 cm intervals, and dated by ^{210}Pb . Limnological trends are summarized by projecting the core diatom assemblages onto the DCCA biplot of the modern samples (Fig. 3 main figure). Most lakes show a net negative trajectory along axis 1 (decreasing pH, increasing DOC), although the pattern varies among lakes. The older sites in the outer coast region (Paps, Harbor Point, Dagelet, Brady) cluster together in the lower left of the plot, whereas the younger sites along the Glacier Bay fjord (Spokane Cove, Blue Mouse, Lester-1, Lester-2, Lester-3), with the exception of Bartlett Lake, cluster in the upper right of the plot. The splitting of the trajectories into two groups suggests that the outer coast sites started at a different point along the gradient of the first axis than did the other lakes. The younger group of lakes also shows an early positive movement along the second axis (increasing TN), whereas the older group of lakes shows the opposite pattern or no net trend.

The core data thus confirm the main trend of lake development suggested by the modern chronosequence: a progressive loss of pH, alkalinity and base cations, and a corresponding increase in DOC. This trend is clearly manifest in all but three of the youngest lakes (Blue Mouse, Spokane, Lester-3), a pattern consistent with chronosequence results showing pH and alkalinity declines beginning some

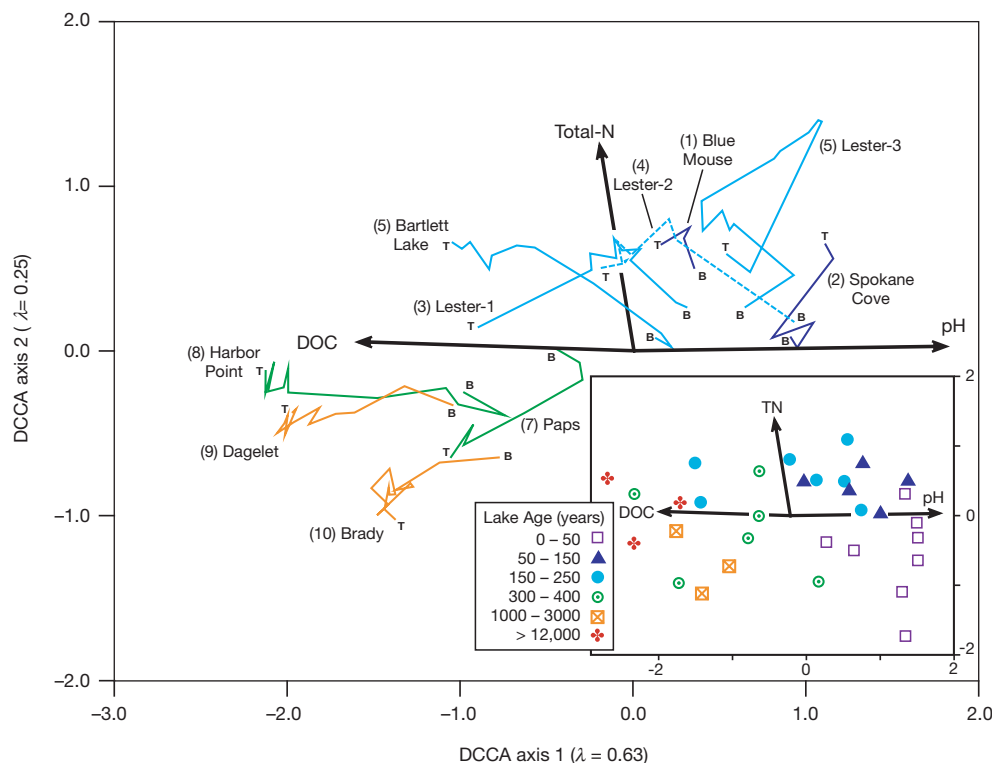


Figure 3 Detrended canonical correspondence analysis of modern and fossil diatom assemblages from the chronosequence lakes. Inset, ordination of modern diatom assemblages (from surface sediments) and associated water-chemistry data; water-chemistry variables were chosen by forward selection (a stepwise addition based on explained variance). The arrows depict the direction of increase and maximum variation of the measured environmental variables; other significant variables (not shown) include Na^+ , Cl^- , SO_4^{2-} and total-Al. Lakes plotting close to one another have similar diatom composition. Main figure, the time-trajectories from 10 sediment cores (B, bottom; T, top)

were projected onto the biplot by adding the core assemblages to the analysis as passive samples; numbers on lake trajectories correspond to map localities in Fig. 1. Diatoms were extracted and identified following standard procedures; a minimum of 500 individual diatoms in surface samples and 400 in core samples were counted. A final dataset of 223 taxa, which included only those with $\geq 0.5\%$ relative abundance in at least one sample from two or more lakes, was used for statistical analysis. Chemical variables that were positively skewed were $\log_{10}$ or square-root transformed before statistical analysis.

200 years after deglaciation. However, the core trajectories indicate that the patterns of change for TN are highly variable and that the TN trend observed in the modern chronosequence cannot be generalized, even to this relatively small geographical region.

Mechanisms of limnological change

The effects of vegetational succession, soil development and hydrologic change, which are all potential agents of lake development, can be inferred by correlation of the limnological trends with those occurring during primary succession in the terrestrial system. In the case of soil development, we have a well documented picture of carbonate loss, pH decline, and carbon and nitrogen increase during the first century of terrestrial succession at Glacier Bay^{24–26}. Thus the weathering of soluble minerals from upper soil horizons may be partially responsible for lakes becoming more dilute and acidic over time, while the steady accretion of soil organic matter probably contributes to a greater flux of organic acids (DOC) that stain the waters of older lakes^{27–29}.

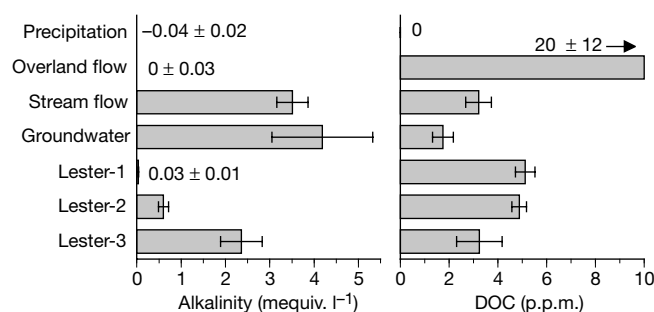


Figure 4 Water chemistry of three Lester Island lakes and their hydrologic sources. Precipitation was sampled monthly by bulk collector over three years; concentrations are volume-weighted means. Stream flows were monitored continuously over a 3-year period by automated stage recorders and V-notched weirs. Groundwater head gradients were measured in littoral areas along the shoreline of each lake with a hydraulic potentiometer (mini-piezometer)⁴². Stream and groundwater chemistry was sampled a total of 3–5 times; values represent combined means from Lester-2 and Lester 3 (Lester-1 lacks inflows); error bars, ± 1 s.d. Overland flow was sampled from temporary pools and forest hollows on Lester Island and in the Bartlett Cover area; values shown are means of 16 collections.

However, soil weathering alone is insufficient to explain the timing of dilution of lake waters in the chronosequence. Lake Ca^{2+} and pH do not decline appreciably over the same period of time that initial soil leaching is thought to occur (0–100 years). Furthermore, while surficial catchment drainage may lose ionic strength because of soil weathering, ground water at depth should be relatively unaffected by these changes. Lakes receiving even a moderate load of dissolved solids from sub-surface flow should not respond to terrestrial succession unless groundwater inputs are reduced in the process. That essentially all of the older sites along the outer coast and on Pleasant Island (most in carbonate-rich terrain) are dilute and at least somewhat acidic implies that groundwater inputs to lakes eventually decrease, either in volume or in concentration.

The most likely cause for a decrease in groundwater flux is a long-term decline in hydrologic recharge associated with soil development. Hardened soil horizons (organic-sesquioxide cemented hardpans) and thick accumulations of peat, which begin forming several hundred years after disturbance (by glaciation or windstorm in southeastern Alaska, gradually inhibit internal soil drainage and increase surface runoff^{26,30,31}. This process should ultimately lower groundwater tables and reduce groundwater discharge to lakes and streams³². Surface runoff (either overland flow or interflow) coursing through peat or weathered soil horizons would bear a much smaller load of dissolved solids and higher concentrations of DOC than the ground water it replaced, and lakes receiving these inputs would gradually become more dilute and acidic.

The development of nitrogen-fixing alder thickets that cause the accretion of soil nitrogen during the middle stages of terrestrial succession is the likely driver for the rise and subsequent decline in lake-water TN depicted in the chronosequence and in core trajectories of younger sites along the Glacier Bay fjord³³. Conversely, the absence of an alder phase in the vegetational development of the outer coast is probably the reason that core trajectories of older lakes do not show this transient rise in nitrogen. Because late Holocene ice advances along the outer coast were of relatively limited extent and deglaciated surfaces were always near spruce seed sources, alder was probably never a significant component of the vegetation^{18–20}. Without alder and the associated nitrogen fixation in soils, older sites along the outer coast did not receive the early pulse of lakewater TN that is recorded by younger lakes in upper Glacier Bay.

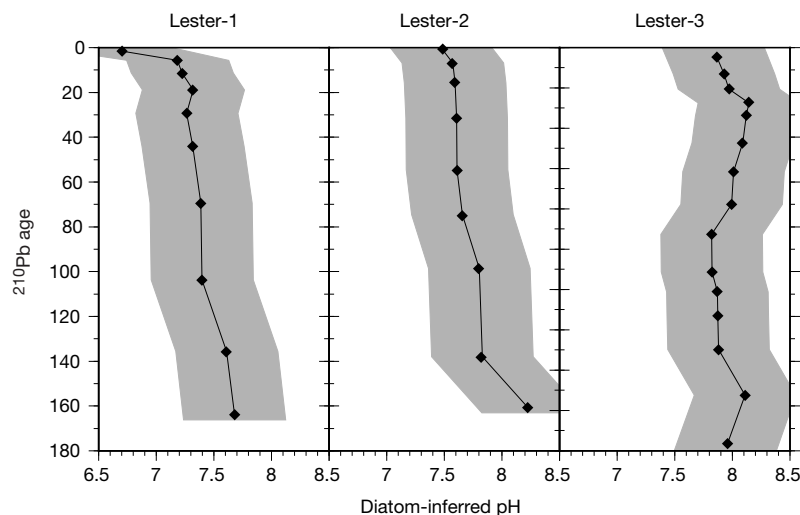


Figure 5 Core trajectories of three hydrologic study lakes on Lester Island. The pH trends were reconstructed from fossil diatom assemblages by using a weighted averaging transfer function⁴³. The shaded regions represent the root mean square error of prediction (0.44 pH units), a robust estimate of the predictive error of the transfer function derived by comparing measured pH and diatom-inferred pH for the suite of modern Glacier Bay lakes

under cross-validation. The pH declines in Lester-1 (1.0) and Lester-2 (0.8) are statistically significant ($P < 0.001$) while the pH change in Lester-3 is not ($P = 0.685$); significance was assessed using linear regression with a restricted Monte Carlo permutation test appropriate for time-series⁴⁴. Ages (years before present) were determined by ^{210}Pb dating.

Geological and hydrologic variability

Although the stratigraphic analyses indicate that a decline in pH over time is a universal trend in our data set, the magnitude and timing of pH change differs, even among lakes of similar age (Fig. 3, main figure). This variability in lake development is likewise reflected in the scatter of modern lake chemistries surrounding the main trends of the chronosequence (Fig. 2). Most solutes show substantial variability among lakes of similar age, while certain sites (for example, Bartlett Lake) are consistent outliers for all ions. In addition to common pathways of lake development, core trajectories and chronosequence results illustrate the individual nature of lake development and its regulation by local factors.

We attribute most of the variability among lake trajectories to differences in hydrologic setting rather than differences in catchment lithology. Because the bedrock geology of Glacier Bay National Park is highly varied³⁴, we initially expected that lake chemistry might vary with localized differences in carbonate content of glacial till. However, we found no discernible relationship between the chemistry of the 11 youngest lakes (<100 years) in the chronosequence and the carbonate content of unweathered soil samples from their catchments³⁵. This point is well illustrated by three hydrologic study sites on Lester Island, which show considerable variability in lake chemistry, even though they are of the same age (200 years), on similar geological substrate, at the same elevation, and in the same stage of vegetational succession. Lake-water alkalinity ranges from 0.03 mequiv. l⁻¹ in Lester-1 to 2.36 mequiv. l⁻¹ in Lester-3, yet the percentage of carbonate in soil parent material ($\leq 62.5 \mu\text{m}$ fraction) from their catchments is very similar (averaging 4–6%).

Differences among the three lakes in their water sources rather than geological substrate account for their present-day differences in chemistry. All of the lakes receive direct precipitation and also some shallow sub-surface flow from the upper soil horizons directly beneath the organic layer (hereafter called overland flow). However, neither precipitation nor overland flow contains significant amounts of alkalinity (Fig. 4), as the upper 50 cm or so of mineral soil on Lester Island is leached of carbonate. Thus the source of the alkalinity must be deeper groundwater flow, which either seeps directly into the lakes or enters them via streamflow. Annual loading of alkalinity via inlet streams ranges from 0 kequiv. in Lester-1 (no stream inflow) to 169 kequiv. in Lester-2 and 302 kequiv. in Lester-3. These inputs are directly proportional to the range of alkalinities in the three lakes (Fig. 4), as are the proportions of lake margin along which positive groundwater head-gradients were measured (0% in Lester-1, 8% in Lester-2, and 45% in Lester-3). Because most groundwater flow in and out of lakes tends to occur near the lake margin³⁶, the length of shoreline with positive head gradient provides a qualitative measure of groundwater in-seepage. Thus the low alkalinity of Lester-1, as well as its high DOC concentration, results from the fact that it is fed exclusively by overland flow (low alkalinity, high DOC) and direct precipitation (low alkalinity, low DOC). By contrast, the hydrologic budget of high-alkalinity Lester-3 is dominated by stream and groundwater inflows (high alkalinity, low DOC), while Lester-2 is intermediate with respect to both lake chemistry and water sources (Fig. 4).

In addition to explaining some of the scatter in the chronosequence data, these hydrologic observations provide a basis for understanding the different pH trajectories of the three Lester Island lakes. Using these hydrologic data, we predicted, *a priori*, that the magnitude of pH decline over the past 200 years should be greatest in Lester-1 and least in Lester-3. This prediction follows from the two suggested mechanisms for alkalinity loss: (1) a relatively rapid leaching of carbonates and other soluble minerals from surface soils, and (2) a relatively slow replacement of groundwater inflow by overland runoff caused by a gradual cementing of deeper soil horizons. The first of these processes is effectively completed within the first 100 years of deglaciation, whereas the second requires several centuries to become manifest^{24–26,30}. Hence,

Lester-1 with little or no groundwater inflow should be most affected by the first process and quickly lose most of its buffering capacity, while Lester-3 with its substantial groundwater supply should show little change within its 200-year life span. Indeed, this is what we found in pH reconstructions from stratigraphic diatom assemblages (Fig. 5): no net change in the pH of Lester-3, and statistically significant pH declines of 0.8 in Lester-2, and 1.0 in Lester-1 ($P < 0.001$). The reconstructions also suggest that Lester-2 and Lester-3 were of approximately equal pH at the time formation, whereas Lester-1 has always been more dilute.

Conclusions

Our comparison of limnological trends, inferred from a chronosequence of lakes and reconstructed from sediment cores, indicates that progressive changes in major-ion chemistry are widespread across broad geographical regions. Patterns of nutrient change, however, are variable over relatively small spatial scales. Even among the more universal trends (decreasing pH), local factors—particularly those related to hydrology—control the initial starting point as well as the rate and magnitude of change. Despite these underlying differences, the Glacier Bay results indicate that lakes may indeed age in a directional manner, driven largely by ecological forces that are set in motion with glacial retreat. Limnological change in the Glacier Bay region appears to be linked biogeochemically to the well studied process of primary terrestrial succession. The rapid re-vegetation of the newly deglaciated landscape^{18–20} and associated changes in the soil—carbonate leaching, nitrogen fixation, humus build-up, and hardpan formation^{24–26,30}—alter the flux of base cations, DOC and nitrogen to downstream lakes, ultimately changing the biological components of the system, including algal communities³⁷, zooplankton^{38,39} and higher aquatic plants³⁵. Clearly, the traditional model of progressive eutrophication over time does not apply to the Glacier Bay study lakes. The cool, moist climate of Alaska's temperate rainforests offers an optimal setting for the soil changes embodied in this model of lake development. However, similar conditions exist throughout the boreal forest regions of North America, Europe and Asia, and so the patterns of limnological change found for Glacier Bay may thus apply to many of the lakes created by continental glaciation. □

Received 25 April; accepted 14 September 2000.

- Rodhe, W. In *Eutrophication: Causes, Consequences, Correctives* (ed. Hutchinson, G. E.) 50–64 (National Academy of Sciences, Washington DC, 1969).
- Pearsall, W. H. The development of vegetation in the English Lakes, considered in relation to the general evolution of glacial lakes and rock basins. *Proc. R. Soc. Lond. B* **92**, 259–284 (1921).
- Naumann, E. Grundzüge der regionalen Limnologie. *Binnengewässer* **11**, 1–176 (1932).
- Thienemann, A. Tropische Seen und Seetypologie. *Arch. Hydrobiol.* **9**, 205–231 (1931).
- Carpenter, S. J. et al. *Nonpoint Pollution of Surface Waters by Phosphorus and Nitrogen* (Ecological Society of America, Washington DC, 1998).
- Whitehead, D. R., Charles, D. F., Jackson, S. T., Smol, J. P. & Engstrom, D. R. The developmental history of Adirondack (N.Y.) lakes. *J. Paleolimnol.* **2**, 185–206 (1989).
- Ford, M. S. A 10,000-yr history of natural ecosystem acidification. *Ecol. Monogr.* **60**, 57–89 (1990).
- Korhola, A. A. & Tikkanen, M. J. Holocene development and early extreme acidification in a small hilltop lake in southern Finland. *Boreas* **20**, 333–356 (1991).
- Renberg, I., Korsman, T. & Anderson, N. J. A temporal perspective of lake acidification in Sweden. *Ambio* **22**, 264–271 (1993).
- Battarbee, R. W., Charles, D. F., Dixit, S. S. & Renberg, I. in *The Diatoms: Applications for the Environmental and Earth Sciences* (eds Stoermer, E. F. & Smol, J. P.) 85–127 (Cambridge Univ. Press, Cambridge, 1999).
- Birks, H. H., Battarbee, R. W. & Birks, H. J. B. The development of the aquatic ecosystem at Kråkenes Lake, western Norway, during the late-glacial and early-Holocene – a synthesis. *J. Paleolimnol.* **23**, 91–114 (2000).
- Matthews, J. A. *The Ecology of Recently Deglaciated Terrain* (Cambridge Univ. Press, Cambridge, 1992).
- Cooper, W. S. The recent ecological history of Glacier Bay, Alaska: II. The present vegetation cycle. *Ecology* **4**, 223–246 (1923).
- Cooper, W. S. A third expedition to Glacier Bay, Alaska. *Ecology* **12**, 61–95 (1931).
- Lawrence, D. B. Recent glacier history of Glacier Bay, Alaska and development of vegetation on deglaciated terrain with special reference to the importance of alder in the succession. *Yb. Am. Phil. Soc.* **1950**, 175–176 (1951).
- Lawrence, D. B. Glaciers and vegetation in southeastern Alaska. *Am. Sci.* **46**, 89–122 (1958).
- Reiners, W. A., Worley, I. A. & Lawrence, D. B. Plant diversity in a chronosequence at Glacier Bay, Alaska. *Ecology* **52**, 55–69 (1971).

18. Chapin, F. S., Walker, L. R., Fastie, C. L. & Sharman, L. C. Mechanisms of primary succession following deglaciation at Glacier Bay, Alaska. *J. Ecol.* **64**, 149–175 (1994).
19. Chapin, F. S., Fastie, C. L., Walker, L. R. & Sharman, L. C. in *Proc. Third Glacier Bay Sci. Symp.* 1993 (ed. Engstrom, D. R.) 96–100 (National Park Service, Anchorage, 1995).
20. Fastie, C. L. Causes and ecosystem consequences of multiple successional pathways of primary succession at Glacier Bay, Alaska. *Ecology* **76**, 1899–1916 (1995).
21. Brower, W. A. Jr, Diaz, H. F., Prechtel, A. S., Searaby, H. W. & Wise, J. L. *Climatic Atlas of the Outer Continental Shelf Waters and Coastal Regions of Alaska Vol. 1, Gulf of Alaska* (Environmental Information and Data Center, University of Alaska, Anchorage, 1977).
22. Hunter, L. E. & Powell, R. D. in *Proc. Third Glacier Bay Sci. Symp.* 1993 (ed. Engstrom, D. R.) 46–54 (National Park Service, Anchorage, 1995).
23. ter Braak, C. J. F. Canonical correspondence analysis: a new eigenvector method for multivariate direct gradient analysis. *Ecology* **67**, 1167–1179 (1986).
24. Crocker, R. L. & Major, J. Soil development in relation to vegetation and surface age at Glacier Bay, Alaska. *J. Ecol.* **43**, 427–448 (1955).
25. Ugolini, F. C. in *Soil Development and Ecological Succession in a Deglaciated Area of Muir Inlet, Southeast Alaska* (ed. Mirsky, A.) 29–72 (Institute of Polar Studies, Ohio State University, Columbus, 1966).
26. Bormann, B. T. & Sidle, R. C. Changes in productivity and distribution of nutrients in a chronosequence at Glacier Bay National Park, Alaska. *J. Ecol.* **78**, 561–578 (1990).
27. Cronan, C. S. & Aiken, G. R. Chemistry and transport of soluble humic substances in forested watersheds of the Adirondack Park, New York. *Geochim. Cosmochim. Acta* **49**, 1697–1705 (1985).
28. Engstrom, D. R. Influence of vegetation and hydrology on the humus budgets of Labrador lakes. *Can. J. Fish. Aquat. Sci.* **44**, 1306–1314 (1987).
29. Rasmussen, J. B., Godbout, L. & Schallenberg, M. The humic content of lake water and its relationship to watershed and lake morphometry. *Limnol. Oceanogr.* **34**, 1336–1343 (1989).
30. Ugolini, F. C. & Mann, D. H. Biopedological origin of peatlands in South East Alaska. *Nature* **281**, 366–368 (1979).
31. Noble, M. G., Lawrence, D. B. & Streveler, G. P. Sphagnum invasion beneath an evergreen forest canopy in southeastern Alaska. *Bryologist* **87**, 119–127 (1984).
32. Almendinger, J. E. in *Proc. Second Glacier Bay Sci. Symp.* (eds Milner, A. M. & Wood, J. D. Jr) 133–135 (National Park Service, Anchorage, Alaska, 1990).
33. Goldman, C. R. The contribution of alder trees (*Alnus tenuifolia*) to the primary productivity of Castle Lake, California. *Ecology* **42**, 282–288 (1961).
34. Brew, D. A. *et al.* Mineral resources of Glacier Bay National Monument wilderness study area, Alaska. *US Geol. Surv. Open-file Rep.* **78-494**, 1–659 (1978).
35. Engstrom, D. R. & Fritz, S. C. in *Proc. Second Glacier Bay Sci. Symp.* (eds Milner, A. M. & Wood, J. D. Jr) 127–132 (National Park Service, Anchorage, 1990).
36. McBride, M. S. & Pfannkuch, H. O. The distribution of seepage within lakebeds. *US Geol. Survey J. Res.* **3**, 505–512 (1975).
37. Olson, O. G. *Mechanisms of Long-term change in Periphytic diatom community structure*. Thesis, Lehigh Univ. (1998).
38. Olson, O. G., Engstrom, D. R. & Fritz, S. C. in *Proc. Third Glacier Bay Sci. Symp.* 1993 (ed. Engstrom, D. R.) 154–163 (National Park Service, Anchorage, 1995).
39. Williamson, C. E. *et al.* Ultraviolet radiation and zooplankton community structure following deglaciation in Glacier Bay, Alaska. *Ecology* (in the press).
40. Mann, D. H. & Ugolini, F. C. Holocene glacial history of the Lituya District, southeast Alaska. *Can. J. Earth Sci.* **22**, 913–928 (1985).
41. Engstrom, D. R., Hansen, B. C. S. & Wright, H. E. Jr A possible Younger Dryas record in southeastern Alaska. *Science* **250**, 1383–1385 (1990).
42. Winter, T. C., LaBaugh, J. W. & Rosenberry, D. O. The design and use of a hydraulic potentiometer for direct measurement of differences in hydraulic head between groundwater and surface water. *Limnol. Oceanogr.* **33**, 1209–1214 (1988).
43. Birks, H. J. B., Line, J. M., Juggins, S., Stevenson, A. C. & terBraak, C. J. F. Diatoms and pH reconstruction. *Phil. Trans. R. Soc. Lond. B* **327**, 263–278 (1990).
44. Juggins, S., Flower, R. J. & Barrarbee, R. W. Palaeolimnological evidence for recent chemical and biological changes in UK Acid Waters Monitoring Network sites. *Freshwat. Biol.* **36**, 203–219 (1996).

Acknowledgements

We thank B.A. Coffin, J.A. Janssens, M.G. Noble, G.O. Seltzer, E.B. Swain and H.E. Wright for assistance in the field; Gustavus residents H. Burd, M. Hervin, R. Howe and G. Streveler for their help; and NPS and USGS staff at Glacier Bay National Park for logistical support. This work was supported by the late D.B. Lawrence and by the National Science Foundation Ecology Program.

Correspondence and requests for materials should be addressed to D.R.E. (e-mail: dengstrom@smm.org).

Structure and assembly of the *Alu* domain of the mammalian signal recognition particle

Oliver Weichenrieder*, Klemens Wild*, Katharina Strub† & Stephen Cusack*

* European Molecular Laboratory Biology, Grenoble Outstation, BP 156X, F-38042, Grenoble Cedex 9, France

† Département de Biologie Cellulaire, Université de Genève, Sciences III, CH-1211, Geneva 4, Switzerland

The *Alu* domain of the mammalian signal recognition particle (SRP) comprises the heterodimer of proteins SRP9 and SRP14 bound to the 5' and 3' terminal sequences of SRP RNA. It retards the ribosomal elongation of signal-peptide-containing proteins before their engagement with the translocation machinery in the endoplasmic reticulum. Here we report two crystal structures of the heterodimer SRP9/14 bound either to the 5' domain or to a construct containing both 5' and 3' domains. We present a model of the complete *Alu* domain that is consistent with extensive biochemical data. SRP9/14 binds strongly to the conserved core of the 5' domain, which forms a U-turn connecting two helical stacks. Reversible docking of the more weakly bound 3' domain might be functionally important in the mechanism of translational regulation. The *Alu* domain structure is probably conserved in other cytoplasmic ribonucleoprotein particles and retroposition intermediates containing SRP9/14-bound RNAs transcribed from *Alu* repeats or related elements in genomic DNA.

The signal recognition particle (SRP) is an essential cytoplasmic ribonucleoprotein particle¹ found in all kingdoms of life that is involved in the targeting of signal-peptide-containing proteins to membranes^{2,3}. In higher eukaryotes, SRP consists of two domains, the *Alu* domain and the S domain, which are linked by the SRP RNA (Fig. 1a). The *Alu* domain comprises the protein heterodimer of SRP9 and SRP14 (denoted SRP9/14) and the 5' and 3' terminal sequences of SRP RNA⁴. This domain is necessary^{5–7} for retarding the ribosomal elongation ('elongation arrest') of nascent chains once their amino-terminal signal sequence is bound by the S domain of SRP. Elongation arrest is thought to enlarge the time window during which SRP can target nascent-chain/ribosome complexes to the membrane of the endoplasmic reticulum, thus ensuring the co-translational mode of protein translocation^{2,8}. Apart from its role in translational control, the *Alu* RNA tertiary structure and probably SRP9/14 binding are crucial in the transcription⁹, maturation¹⁰, nucleolus localization and transport¹¹ of SRP RNA.

The *Alu* RNA tertiary structure is also relevant to understanding the spread of repetitive *Alu* elements, which now account for some 10% of primate genomic DNA (16.8% of human chromosome 22; ref. 12). *Alu* elements probably originated from the SRP RNA gene and propagate through retroposition¹³, an ongoing process that has had a significant impact on the evolution of the human genome¹⁴. Transcriptionally active dimeric *Alu* elements give rise to stable small cytoplasmic *Alu* ribonucleoprotein particles (scAluRNPs) containing bound SRP9/14 (refs 15, 16). A cellular function of such scAlu RNPs remains to be identified. A related example is the neuron-specific and regulated expression in primates of the monomeric *Alu*-like BC200 RNA, which binds SRP9/14 *in vitro*¹⁷ and *in vivo*¹⁸.

We previously determined the structure of a murine SRP9/14 fusion protein, which showed that the heterodimer has remarkable pseudo-two-fold symmetry as a consequence of the similar (β) α β β α fold of both proteins. The four α -helices are packed on the convex side of the continuous and highly concave central six-stranded, antiparallel β -sheet¹⁹. We also defined a minimal *Alu* RNA folding domain²⁰, SA86 (Fig. 1b), which specifically binds SRP9/14 with affinity comparable to that of complete SRP RNA.

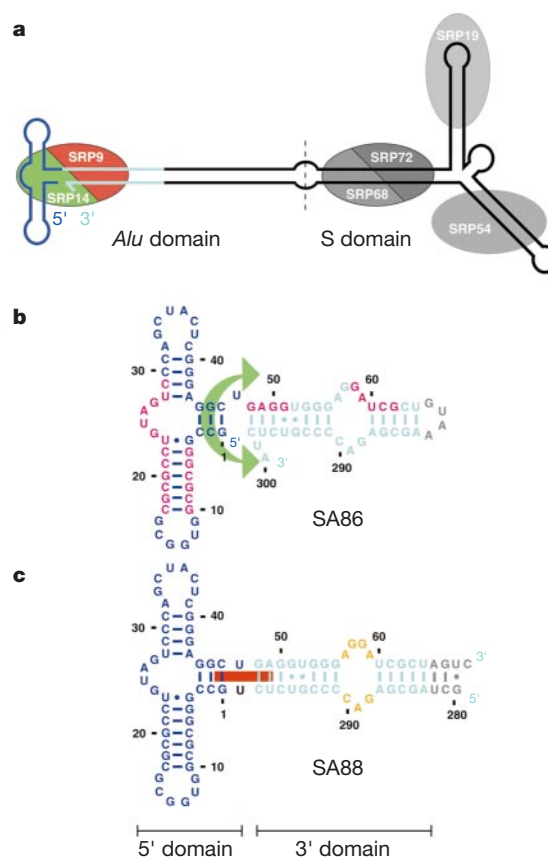


Figure 1 Mammalian SRP and *Alu* domain RNA constructs. **a**, Mammalian SRP with SRP RNA bound to six proteins. The *Alu* domain is coloured red (SRP9), green (SRP14), blue (*Alu* RNA 5' domain) and cyan (*Alu* RNA 3' domain). **b**, SA86, the minimal *Alu* RNA folding domain. S-domain RNA is replaced by a GUAA tetraloop (grey). The green arrow indicates flexibility between the RNA 5' and 3' domains. Hydroxyl-radical footprints of SRP9/14 on SRP RNA are highlighted in magenta²¹. **c**, SA88, a circular permutation of SA86. The original 5' and 3' ends are connected by a mono-uridine linker (black), constraining the RNA 5' and 3' domains to stack in an extended conformation (red bar). The GUAA tetraloop is replaced by a terminal stem (grey). The asymmetric internal loop is highlighted in yellow.

SA86 can be separated into a 5' domain, comprising the first 47 nucleotides of SRP RNA, and a 39-nucleotide 3' domain, comprising a helical stem with an asymmetric internal loop (Fig. 1). The *Alu* RNA 5' domain, which contains the most highly conserved region of *Alu* RNA²¹, is the primary SRP9/14 binding target and the 3' stem contributes significantly to protein affinity only if covalently linked in a flexible way (unpublished data). This result prompted us to crystallize the ternary complex of human SRP9/14 with a 50-nucleotide 5' domain RNA, denoted SA50 (Fig. 2a). We subsequently solved the structure of SRP9/14 in complex with a larger RNA, from which we have deduced a model for the native *Alu* RNP that is consistent with extensive biochemical data.

Structure of the 5' domain complex

Microcrystals of the SA50 *Alu* RNP diffracted X-rays to 3.0 Å resolution at the microfocus beamline²² of the European Synchrotron Radiation Facility (see Supplementary Information). The structure was solved by molecular replacement using the murine SRP9/14 fusion protein¹⁹ as a model. The asymmetric unit contains a specific SRP9/14–SA50 complex with an additional SRP9/14 heterodimer nonspecifically bound to the RNA. Despite the modest resolution (3.2 Å), there is no ambiguity in the RNA fold (Fig. 2b).

The *Alu* RNA core contains a three-way τ -junction. The compactly folded RNA consists of two helical stacks connected by a central U-turn (Fig. 2b, d)—a tertiary structural motif that we call the ' τ -junction' owing to its shape. This fold is common to two other three-way junction RNAs, the hammerhead ribozyme^{23,24} and the 23S ribosomal RNA-binding site of ribosomal protein L11 (L11 RNA)^{25,26}. In all three structures (Fig. 2d), the two helical stacks (H1

and H2) are formed by three RNA strands, S1, S2 and S3. Strand S3 belongs entirely to helix H1 and forms base pairs with both strands S1 and S2, thereby defining helices H1.1 and H1.2. Strand S1 switches from helix H1.1 to H2, bending strongly at the junction, and strand S2 switches from helix H2 to H1.2, with a sharp loop in between containing the characteristic U-turn.

In SA50, strand S1 (nucleotides –2 to 10) is base paired throughout, bending and switching helices between C3 and G4 (Fig. 2). Strand S3 (nucleotides 40–48) is also completely base paired, and adopts helical geometry. The central U-turn is formed by U25 stacking on G24 and hydrogen bonding to the U28 phosphate. This reverses the chain direction of strand S2 (nucleotides 17–31), constraining the relative orientation of helices H2 and H1.2. The unpaired G24 makes a cross-strand purine stack on G4 of the opposing strand S1. G4·U23 is a universally conserved wobble base pair, which terminates helix H2. The conserved purine bases A26 and G27 loop out of the U-turn and are free for potential interactions.

The *Alu* RNA stem-loops make tertiary interactions. The relative orientation of subdomain 1 (helices H1.1 and H1.2 with loop L1.2) to subdomain 2 (helix H2 with loop L2) is reinforced by tertiary interactions between loops L1.2 and L2 (Fig. 2c). Loop L2 is a rigid hexa-loop that is internally stabilized by a sheared G11·G16 pair and a U-turn at U12, a structural motif similar to the 1095-loop of L11 RNA^{25,26}. Bases G13, G14 and C15 face outward and form antiparallel stacked Watson–Crick pairs with C37, C34 and G33 of loop L1.2. Loop L1.2 has an unusual conformation, which may be induced only on interaction with loop L2 as the absence of internal stabilizing interactions suggests that it may be somewhat plastic. It is sharply bent with respect to helix H1.2, with the unpaired base A32 of helix H1.2 being almost perpendicular to G33 in loop L1.2. The

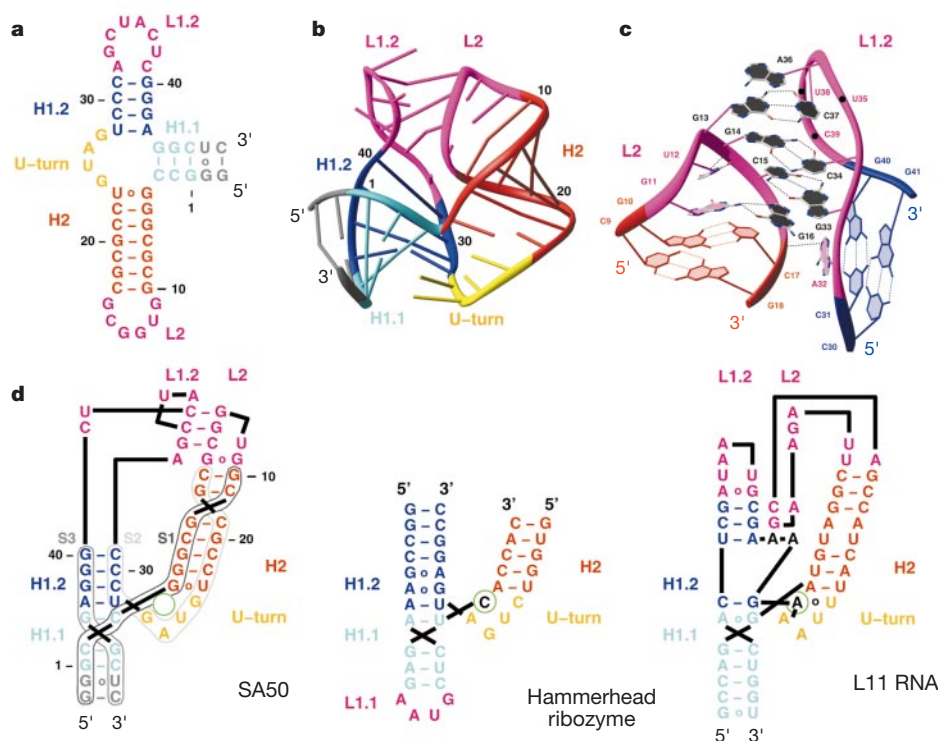


Figure 2 Structure of the *Alu* RNA 5' domain (SA50). **a**, Secondary structure with the two artificial base pairs in grey. H, helix; L, loop; S, strand. **b**, Three-dimensional ribbon representation. **c**, Close-up of the interaction of the two peripheral loops. The view from the back emphasizes the central sandwich of stacked bases (black). U35, U38 and C39 are omitted for clarity. Dotted lines indicate putative hydrogen bonds. **d**, The τ -junction, a three-way junction comprising two helical stacks connected by a U-turn. Two-

dimensional topology diagrams derived from the crystal structures of SA50, the hammerhead ribozyme^{23,24} and L11 RNA (a fragment of 23S rRNA)^{25,26} show the common architecture of the cores and the different interactions of the peripheral loops. An extra nucleotide in strand S1 of the ribozyme and in strand S2 of L11 RNA is absent in SA50 (green circle).

three interloop base pairs are sandwiched between two lids, G16 at the bottom and A36 at the top. U35, U38 and C39 in loop L1.2 are solvent exposed.

SRP9/14 specifically recognizes the core of the τ -junction. In the RNP, SRP9/14 is bound on the opposite side of the RNA with respect to the interacting loops (Fig. 3a). The human SRP9/14 heterodimer closely resembles the murine fusion protein, and RNA complexation does not change the backbone significantly. The N- and C-terminal residues of SRP9 (residues 1–5 and 76–86) and the C-terminal residues of SRP14 (amino acids 96–107) are disordered, and the internal loop of SRP14 (residues 36–53) is untraceable. The centres of mass of SRP9/14 and SA50 are well separated either side of the large interface (1,500 Å² buried surface area), and the positively charged concave β -sheet surface is the primary RNA-binding site. The protein interacts specifically with the core of the τ -junction, contacting both helical subdomains and the U-turn exclusively on RNA strand S2. Despite its pseudosymmetry, SRP9/14 is bound asymmetrically: the bulk of SA50 leans towards the SRP14 side of the heterodimer, with no contacts to the β 2 and β 3 strands of SRP9. The protein contacts RNA principally by basic residues to the RNA backbone rather than through specific hydrogen bonds to bases, and there are no base-stacking interactions with aromatic residues.

The backbone of RNA strand S2 follows a curved canyon formed mostly by basic amino acids, providing a complementary shape and charge distribution (Fig. 3b). Putative hydrogen bonds, often multiple, are made to every phosphate from C19 to C29, except

G20 (Fig. 3c). Between nucleotides C21 and U25, the backbone of RNA strand S2 parallels the SRP9–SRP14 interface. Residues from the β 1 strands of both SRP9 and SRP14 contact phosphates C22 to A26, highlighting the need for protein heterodimerization as a prerequisite for RNA binding²⁷.

Around the U-turn, the RNA backbone runs perpendicular to the β -sheet, such that nucleotides G24 to C29 contact residues from five different β -strands, including the short N-terminal β -strand of SRP14 (Fig. 3c). Phosphates G24 to G27 are embedded in a dense hydrogen-bond network with basic side chains typically bridging adjacent phosphates. Of note is the highly conserved SRP14 Arg 59, which bridges phosphates U25 and G27, thus recognizing the particular backbone conformation of the U-turn. G24 and U25, key residues within the τ -junction, are the most extensively contacted and the only base-specifically recognized nucleotides. The centrally positioned SRP14 Lys 66 bridges O6 of G24 and O4 of U25 and, in addition, contacts the U28 phosphate. This reinforces the fold of SA50 by doubling the crucial U-turn hydrogen bond to the U28 phosphate. Topologically, SRP14 Lys 66 occupies the position of the extra nucleotide present in the τ -junctions of the hammerhead ribozyme and of L11 RNA (Fig. 2d). The position of RNA helix H1.2 is also maintained by the SRP14 N-terminus (SRP14 Val 2) which contacts the C29 phosphate. The role of SRP14 Lys 64 (which forms a remarkable triskelion with Lys 65 and Lys 66) in pinning a distant region of helix H2 (C19 phosphate) to the central core of the complex is of particular note (Fig. 3c).

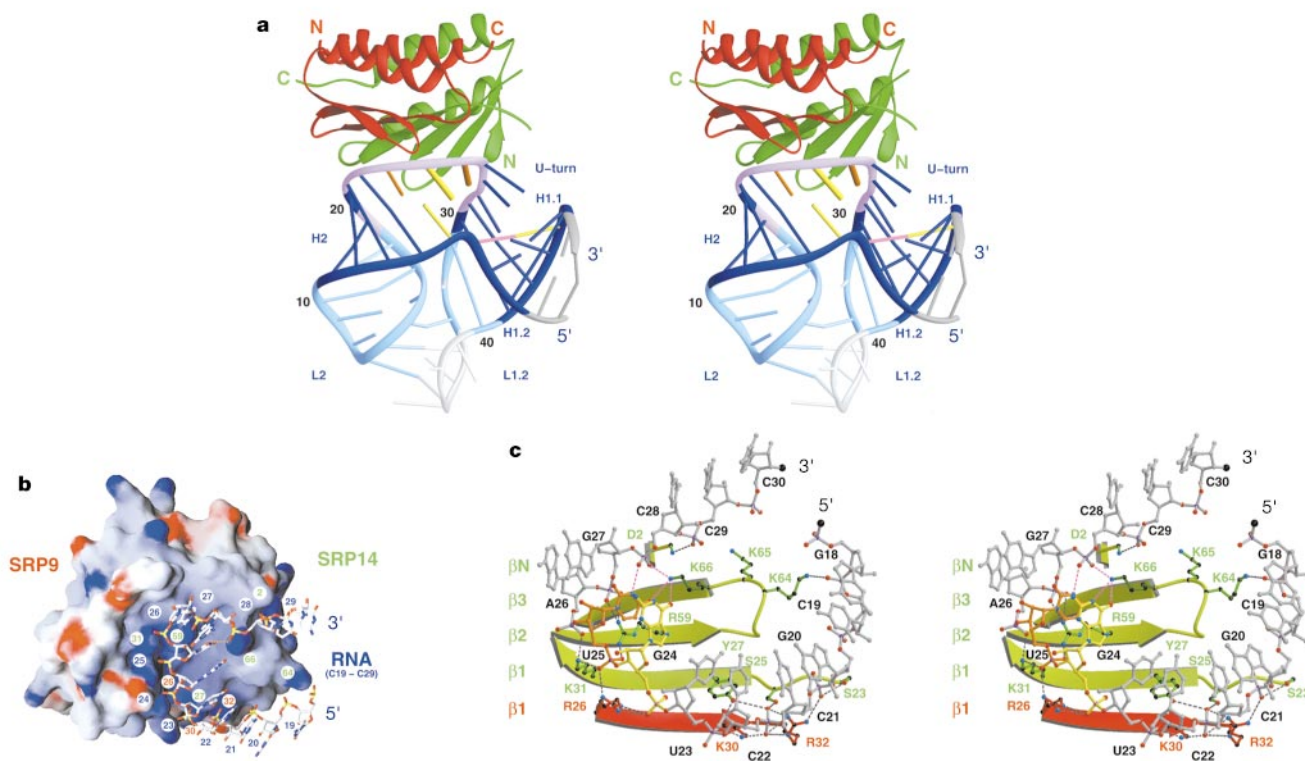


Figure 3 Protein recognition of the conserved fold of the *Alu* RNA τ -junction. **a**, Stereo view of the ternary complex of SRP9 (red), SRP14 (green) and the *Alu* RNA 5' domain (SA50, blue with the peripheral loops pointing down and the two artificial base pairs in grey). Bases highly conserved in all *Alu*-like RNAs are coloured orange (G), magenta (C) or yellow (U); conserved base pairs and U-turn are dark blue; nucleotides occasionally absent even in higher eukaryotes (for example, plants, nematodes, insects) are silver; the RNA backbone contacted by protein is violet. Thick cylinders indicate RNA bases specifically recognized by protein. **b**, The GRASP⁴⁸ surface potential of SRP9/14 with RNA strand S2 in ball-and-stick representation shows the complementarity of shape and charge. The

RNA backbone (phosphates numbered in blue) follows a canyon formed by basic residues from both SRP9 and SRP14 (red and green numbers, respectively). The U-turn hydrogen bond is dotted orange. See **c** for residue identity. **c**, Stereo diagram (view similar to **b**) of the RNA–protein interaction with part of the protein β -sheet in a ribbon representation and putative hydrogen bonds dotted. Important protein side chains and RNA strand S2 are in ball-and-stick representation. G24 (yellow) and U25 (orange) are base-specifically recognized. Interactions contributing to the positioning of phosphate U28 are highlighted in magenta.

Structure of SRP9/14 complex with 88-nucleotide RNA

To extend our structural understanding to the complete *Alu* domain, we crystallized purified *Alu* RNPs made from human SRP9/14 and SA86, the minimal *Alu* RNA, or one of two circularly permuted RNA variants. In SA91, which mimics the 5′–3′ domain flexibility of wild-type RNA, the natural 5′ and 3′ ends are connected by four uridines (not shown), whereas in SA88 (Fig. 1c) a rigidifying linker of one uridine is used. SA88 RNP crystals grown in the presence of europium nitrate diffracted beyond 4 Å resolution, but anisotropically. A good, unbiased, 4 Å resolution experimental map (Fig. 4) was obtained by the method of multiple anomalous dispersion (MAD) using the europium L(III) edge, and a complete model with an *R* factor of 38.9% was constructed, lacking only the last 2 of the 88 nucleotides (see Supplementary Information).

The structure of the SA88 RNP (Fig. 5a) shows the protein and RNA 5′ domain to be essentially the same as in the SA50 RNP, apart from a slight variation in RNA loop L1.2 probably caused by crystal contacts. As expected, the *Alu* RNA 3′ domain of SA88 (denoted helix H3) stacks onto helix H1.1 forming an extended helical structure. The proximal (helix H3.1) and distal (helix H3.3) parts of helix H3 have rather regular, A-helical parameters. The central part (helix H3.2), corresponding to the asymmetric internal loop in the secondary structure, has a unique geometry that causes considerable widening of the major groove and bends the helical axis between helices H3.1 and H3.3.

A model for the native *Alu* domain

The SA88 RNP forms a crystallographic dimer in which SRP9/14 clamps within its concave β-sheet surface not only the conserved region of the RNA 5′ domain, but also the 3′ domain of a two-fold related neighbouring molecule (Fig. 5b). The RNP dimer results entirely from the dual recognition by each SRP9/14 of both 5′ and 3′ domains from different RNA molecules, there being no contact between the two SRP9/14 molecules. By a number of analytic techniques, we have found that in concentrated solution the SA88 RNP is dimeric, whereas the SA86 and SA91 RNPs are monomeric, as are the three free RNAs (unpublished results). We therefore

propose that intermolecular interactions leading to dimerization allow the rigidly constrained SA88 RNP to mimic intramolecular interactions in the SA86 and SA91 RNPs (and native SRP), where the 3′ domain is linked flexibly to the 5′ domain.

We generated a model for the monomeric *Alu* RNP by adjusting solely the position of U47 to cross-connect its ribose O3′ to the phosphate of G48 of the neighbouring molecule in the crystallographic dimer, 9 Å away. This places the two domains of a single RNA molecule side by side in the concave β-sheet surface of a single SRP9/14 heterodimer, the 3′ domain occupying the SRP9 surface left vacant in the SA50 RNP structure (Fig. 5d). Thus, we propose that SRP9/14 binding induces the RNA to fold back on itself like a jack-knife, such that the stem leading to the SRP S domain emerges on the SRP14 C-terminal side of SRP9/14 (Fig. 5b, c). Most of the 3′ domain contacts are made to the minor groove of distorted helix H3.2 (the asymmetric internal ‘loop’). The β2–β3 loop of SRP9 is inserted into it, and the minor groove side of helix H2 of the RNA 5′ domain packs against it. The implied back-folding of the RNA 3′ domain by 180° is probably the final, but reversible, step in the assembly pathway of the *Alu* domain (unpublished data).

The *Alu* domain model and biochemical data

Extensive biochemical data are explained well by the *Alu* domain model. Most notably, it can rationalize the complex footprint of SRP9/14 on *Alu* RNA in hydroxyl-radical cleavage experiments²¹. Whereas the SA50 RNP structure explains the protection seen on strand S2 (owing to direct binding of the protein to the RNA backbone), protection within the 5′ and 3′ domains of SRP *Alu* RNA can be explained only by the model of the compactly folded monomeric *Alu* RNP (Figs 1b, 5c and d). Protected regions I (G5–G10, defined in ref. 21), and III (G48–G51) appear to arise exclusively from protein-induced RNA tertiary contacts, whereas regions II (C17–C29) and IV (G58–G62) are caused by a combination of both induced RNA tertiary contacts and direct protein contacts. Furthermore, only in the compact particle are SRP9 Cys 39 and SRP9 Cys 48 buried in the RNA–protein interface (Fig. 5d), which explains their protection from alkylation in the presence of

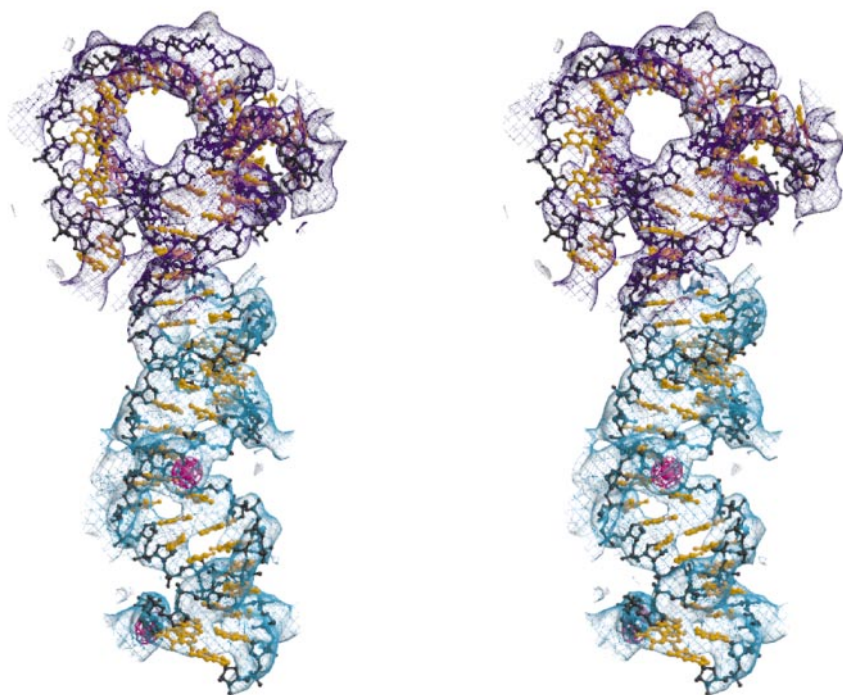


Figure 4 Stereo image of experimental electron density for SA88 RNA. The map around the model of SA88 *Alu* RNA (phosphate-ribose backbone in black, bases in orange) is

contoured in blue (5′ domain) and cyan (3′ domain) at 1.0σ. The central (major) and peripheral (minor) europium sites are contoured in magenta at 8.0 and 3.0σ, respectively.

SRP RNA²⁸. In addition, the model explains the need of protein hetero-dimerization for RNA binding²⁷, the phylogenetically supported RNA base-pairing scheme including tertiary interactions²⁹, the importance of G24 for protein binding³⁰ and the stacking of helix H1.1 with H1.2, as predicted from the RNase V1 cleavage pattern²⁰. The interaction of the SRP14 Val2 N-terminus with phosphate C29 is consistent with the observed absence of SRP14 Met 1 *in vivo*³¹. The model does not yet address the role of the disordered loop between the β 1 and β 2 strands of SRP14, which may order on interaction with the compact, monomeric *Alu* RNA in the region of the bend (Fig. 5a). Finally, the dimensions of the model ($52 \times 59 \times 70 \text{ \AA}^3$) correspond well to the small end domain of SRP observed in electron microscopy³². We therefore conclude that the compactly folded *Alu* RNP model is a close approximation to the true conformation of the SRP *Alu* domain.

Biological implications

We have shown that SRP9/14 binds primarily to the universally conserved core of the *Alu* RNA 5' domain, which forms a U-turn in the context of a τ -junction. This RNA motif is highly conserved in SRP RNAs from higher eukaryotes to yeast and from archaea to some Gram-positive eubacteria²¹. Also conserved is the potential of the RNA 5' domain to back-fold onto the RNA 3' domain. The more peripheral parts of the RNA not involved in protein binding

are much less conserved. Sequence variations or truncations in loop L1.2, for example, indicate that the RNA loop-loop tertiary interactions are likely to be idiosyncratic. With respect to SRP9/14, no eubacterial or archaeal sequence homologues have been found, but the histone-like protein HBSu from *Bacillus subtilis* might be a functional analogue in eubacteria³³. In yeast, where elongation arrest is preserved⁷ in the context of an apparently severely truncated *Alu* RNA, SRP9/14 is replaced by a homodimer of SRP14 (ref. 34). These variations on the basic architecture of the *Alu* domain, together with the structural information on the mammalian system, should help pinpoint features that are maintained owing to functional importance other than *Alu* RNP assembly. The conserved C-terminal tail of SRP14 might represent such an essential feature, as its removal specifically abolishes elongation arrest activity in both the mammalian⁶ and yeast⁷ systems.

Our structural results lead to three conclusions. First, the existence of a stable 5' domain RNP shows the possibility for the *Alu* RNA 5' domain to form a specific complex with nuclear SRP9/14 immediately after transcription by RNA polymerase III. This is consistent with mutational data⁹, which show that the integrity of the tertiary structure and the SRP9/14 binding determinants of the first 50 nucleotides of SRP RNA are necessary for its efficient transcription. SRP9/14 and SRP RNA might therefore associate into a functional unit at the very beginning of transcription, and probably remain tightly associated throughout subsequent RNA processing, assembly and transport steps as well as in the functional cytoplasmic SRP.

Second, regarding the mechanism of elongation arrest, the flexible, rod-like appearance of SRP with its two functional domains at opposing ends has prompted speculations that SRP might reach from near the ribosome polypeptide exit site to the elongation factor/transfer-RNA-binding site^{32,35}. Elongation arrest might then arise from steric interference or active competition of the *Alu* domain with elongation factor (EF)-1 α , EF2 or tRNA. Although the *Alu* domain model shows no structural mimicry with tRNA or elongation factors, its size and shape do not preclude it from entering the ribosomal elongation-factor-binding site. Our hypothesis that the fairly weak interaction of the RNA 3' domain with the stable RNP 5' domain is dynamic suggests that a reversible switch in the folding of the *Alu* domain may also be of functional importance, for example, permitting alternating onset or release of elongation arrest. Further investigation of the mechanism of elongation arrest will require structural work on ribosome-SRP complexes and specific crosslinks to be found between the *Alu* domain and ribosomal components.

Third, RNA sequence analysis in the light of the crystal structure illustrates that non-SRP *Alu* RNAs, such as monomeric BC200 RNA and left and right arm monomers of actively retroposing human *Alu* element RNA, have indeed conserved both the three-dimensional architecture found in SRP *Alu* RNA and the determinants for binding SRP9/14. The *Alu* RNP 5'–3' domain flexibility proposed for SRP might therefore also be important for the processing and function of sc*Alu* RNPs. Furthermore, the observation that in a crystalline environment even SA86 *Alu* RNPs can form the domain-switch non-covalent dimers observed for SA88 RNPs (our own unpublished data) may be of relevance to *Alu* retroposition. Whereas there is no rationale to support the occurrence of dimerization of SRP, a dimeric *Alu* RNP complex might be important in the origin or propagation of tandemly arranged *Alu* retroposons, as retropositional success was clearly correlated with the emergence of dimeric *Alu* elements during primate evolution³⁶.

Conclusions

We have described the structure of two RNPs derived from the mammalian SRP permitting us to propose a model for the *Alu* domain occurring in SRP or other scRNPs. The structures provide further insight into the principles determining the architecture of

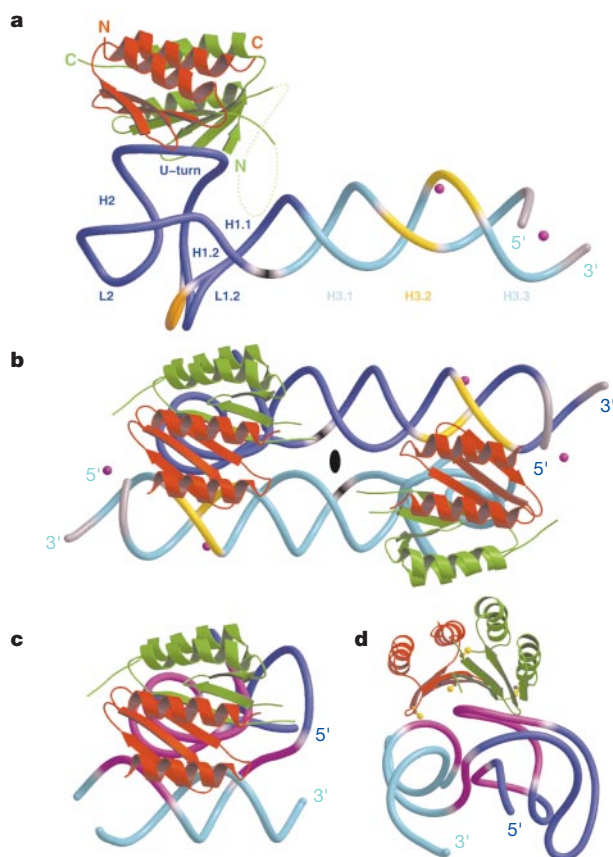


Figure 5 Structure of the *Alu* domain of the mammalian SRP (see Fig. 1 for colours). **a**, The SA88 *Alu* RNP with the alternative conformation of loop L1.2 in the SA50 *Alu* RNP in orange and the two europium sites in magenta. The disordered SRP14 β 1– β 2 loop is indicated by a dotted line; helix H3.2 is coloured yellow; H, helix; L, loop. **b**, The SA88 *Alu* RNP dimer viewed down its crystallographic two-fold axis. **c**, Model for an *Alu* RNP (SA86) in its fully folded, physiological conformation with hydroxyl-radical footprints of SRP9/14 on SRP RNA²¹ in magenta. **d**, As **c**, viewed down the SRP9/14 β -sheet and illustrating the coverage of the previously exposed SRP9 surface with cysteines in yellow ball-and-stick representation.

compactly folded RNAs and the hierarchical assembly of RNPs. They also provide a solid structural basis to assist further functional studies aimed at understanding the role of the *Alu* domain of SRP and possibly other *Alu*-like particles, such as the BC200 RNP, in translational regulation. The finding that SRP9/14 binds specifically to the 5' domain of SRP RNA gives a possible rationalization for the required integrity of this domain in the early steps of SRP RNA transcription. Finally, the structure will contribute to an understanding of the remarkable success of *Alu* retrotransposons in colonizing primate genomes. □

Methods

Preparation of RNA and protein samples

RNA molecules with 3'-terminal hammerhead ribozymes were synthesized by *in vitro* transcription with T7 RNA polymerase from *Hind*III linearized plasmid templates and purified as described²⁰. Plasmids pSA86(2)H, pSA88(2)H and pSA50H, coding for SA86, SA88 and SA50, were obtained by introducing synthetic oligonucleotides into *Aval*/XbaI digested plasmids pSA86H and pSA50 or *Xba*I/*Hind*III digested plasmids pSA88H and pSA85 (ref. 20). Human SRP9 and C-terminally truncated SRP14 were overexpressed in *Escherichia coli* from plasmids pEh9 and pEh14ΔR (ref. 17), separately purified by Heparin and MonoS ion exchange chromatography and finally combined and repurified on a MonoS column as an essentially RNase-free heterodimer.

Crystallization, structure determination and refinement

For the SA50 complex, crystals ($30 \times 5 \times 5 \mu\text{m}^3$) were grown at 12 °C in 3 d in a 3- μl hanging drop (130 μM SA50, 180 μM SRP9/14, 2.5 mM HEPES pH 7.5, 25 mM NaCH₃COO pH 5.0, 10 mM MgCl₂, 75 mM NaCl, 210 mM (NH₄)₂SO₄, 5 mM dithiothreitol, 11% PEG 2000) equilibrated over a 500- μl reservoir (50 mM NaCH₃COO pH 5.0, 10 mM MgCl₂, 140 mM NaCl, 390 mM (NH₄)₂SO₄, 21% PEG 2000). They were flash frozen in liquid nitrogen from 40 mM NaCH₃COO pH 5.0, 15 mM MgCl₂, 120 mM NaCl, 350 mM (NH₄)₂SO₄, 22% PEG 2000 supplemented with 20% glycerol. Crystals are of space group C22₂ with unit-cell dimensions $a = 57.4 \text{ \AA}$, $b = 186.6 \text{ \AA}$ and $c = 189.8 \text{ \AA}$, and 71% solvent. Diffraction data were collected on ESRF beamline ID13 from a single crystal centred in a 10- μm beam. This dataset was integrated with MOSFLM 6.0 (ref. 37) and completed with low-resolution data (50 $\text{\AA}$ to 8.5 $\text{\AA}$) from a second crystal measured at beamline ID14eh2. Molecular replacement was done with AMoRe³⁸ using data between 15 and 5.5 $\text{\AA}$ resolution from the second crystal and searching for two copies of a modified polyaniline model of murine SRP9/14. The top solution was rigid-body refined in CNS³⁹. Two rounds of density modification by program ARP/wARP⁴⁰ yielded the electron density map into which the model was built manually with the program O⁴¹. Refinement (including atomic B factors) was continued with CNS, correcting for anisotropy of the data, applying a bulk solvent correction and restraining identified Watson–Crick and G–U wobble base pairs as well as the ribose sugar pucker (all C3' endo except nucleotides 128, 132, 135–136 and 138–139, which are C2' endo).

For the SA88 complex, crystals ($200 \times 200 \times 150 \mu\text{m}^3$) were grown at 12 °C in a week in 3- μl hanging drops (70 μM SA88 RNP, 28 mM HEPES pH 7.0, 10 mM MgCl₂, 75 mM NaCl, 0.40 mM Eu(NO₃)₃, 5 mM dithiothreitol, 210 mM (NH₄)₂SO₄, 12% PEG 400) equilibrated over 500 μl reservoirs (50 mM HEPES pH 7.0, 10 mM MgCl₂, 150 mM NaCl, 0.80 mM Eu(NO₃)₃, 390 mM (NH₄)₂SO₄, 23% PEG 400). They were flash frozen from mother liquor in liquid nitrogen. Crystals are of space group $P4_22_2$ with unit-cell dimensions of $a = b = 143.3 \text{ \AA}$ and $c = 60.4 \text{ \AA}$, and 67% solvent. Electron density calculated from a molecular replacement solution with the SA50 RNP showed a specific europium-binding site. A three-wavelength MAD experiment at the europium L(III) edge was therefore undertaken on a single crystal at ESRF beamline BM14 and data reduced using the HKL package⁴². A dataset to 4 $\text{\AA}$ resolution was collected on the same crystal at beamline ID14eh2 and processed with MOSFLM 6.0 (ref. 37). SHARP⁴³ was used with the MAD data to identify a second europium-binding site, to refine the heavy metal positions and to calculate phases to 4 $\text{\AA}$, followed by solvent flipping with SOLOMON⁴⁴. The helical *Alu* RNA 3' domain extending the structure of the SA50 RNP was built manually from RNA fragments of known structure. Refinement was done with CNS, applying a correction for data anisotropy ($B_{11} = -6.0 \text{ \AA}^2$, $B_{22} = -6.0 \text{ \AA}^2$, $B_{33} = 12.0 \text{ \AA}^2$), a bulk-solvent correction, rigid-body refinement and a final, highly restrained energy minimization step to improve the RNA geometry.

Figures were produced with RIBBONS⁴⁵ and BobScript⁴⁶/Raster 3D⁴⁷.

Received 11 May; accepted 4 October 2000.

- Walter, P. & Blobel, G. Signal recognition particle contains a 7S RNA essential for protein translocation across the endoplasmic reticulum. *Nature* **299**, 691–698 (1982).
- Walter, P. & Johnson, A. E. Signal sequence recognition and protein targeting to the endoplasmic reticulum membrane. *Annu. Rev. Cell Biol.* **10**, 87–119 (1994).
- Lütcke, H. Signal recognition particle (SRP), a ubiquitous initiator of protein translocation. *Eur. J. Biochem.* **228**, 531–550 (1995).
- Gundelfinger, E. D., Krause, E., Melli, M. & Dobberstein, B. The organization of the 7SL RNA in the signal recognition particle. *Nucleic Acids Res.* **11**, 7363–7374 (1983).
- Siegel, V. & Walter, P. Removal of the *Alu* structural domain from signal recognition particle leaves its protein translocation activity intact. *Nature* **320**, 81–84 (1986).
- Thomas, Y., Bui, N. & Strub, K. A truncation in the 14 kDa protein of the signal recognition particle leads to tertiary structure changes in the RNA and abolishes the elongation arrest activity of the

particle. *Nucleic Acids Res.* **25**, 1920–1929 (1997).

- Mason, N., Ciuffo, L. F. & Brown, J. D. Elongation arrest is a physiologically important function of signal recognition particle. *EMBO J.* **19**, 4164–4174 (2000).
- Rapoport, T. A., Heinrich, R., Walter, P. & Schulmeister, T. Mathematical modeling of the effects of the signal recognition particle on translation and translocation of proteins across the endoplasmic reticulum membrane. *J. Mol. Biol.* **195**, 621–636 (1987).
- Emde, G., Frontzek, A. & Benecke, B. J. Secondary structure of the nascent 7S L RNA mediates efficient transcription by RNA polymerase III. *RNA* **3**, 538–549 (1997).
- Chen, Y., Sinha, K., Perumal, K., Gu, J. & Reddy, R. Accurate 3' end processing and adenylation of human signal recognition particle RNA and *Alu* RNA *in vitro*. *J. Biol. Chem.* **273**, 35023–35031 (1998).
- Jacobson, M. R. & Pederson, T. Localization of signal recognition particle RNA in the nucleolus of mammalian cells. *Proc. Natl Acad. Sci. USA* **95**, 7981–7986 (1998).
- Dunham, I. *et al.* The DNA sequence of human chromosome 22. *Nature* **402**, 489–495 (1999).
- Weiner, A. M., Deininger, P. L. & Efstratiadis, A. Nonviral retrotransposons: genes, pseudogenes, and transposable elements generated by the reverse flow of genetic information. *Annu. Rev. Biochem.* **55**, 631–661 (1986).
- Kazazian, H. H. Jr Mobile elements and disease. *Curr. Opin. Genet. Dev.* **8**, 343–350 (1998).
- Bovia, F., Fornallaz, M., Leffers, H. & Strub, K. The SRP9/14 subunit of the signal recognition particle (SRP) is present in more than 20-fold excess over SRP in primate cells and exists primarily free but also in complex with small cytoplasmic *Alu* RNAs. *Mol. Biol. Cell* **6**, 471–484 (1995).
- Chang, D. Y., Hsu, K. & Marais, R. J. Monomeric scAlu and nascent dimeric *Alu* RNAs induced by adenovirus are assembled into SRP9/14-containing RNPs in HeLa cells. *Nucleic Acids Res.* **24**, 4165–4170 (1996).
- Bovia, F., Wolff, N., Ryser, S. & Strub, K. The SRP9/14 subunit of the human signal recognition particle binds to a variety of *Alu*-like RNAs and with higher affinity than its mouse homolog. *Nucleic Acids Res.* **25**, 318–326 (1997).
- Kremerkothen, J. *et al.* Heterodimer SRP9/14 is an integral part of the neural BC200 RNP in primate brain. *Neurosci. Lett.* **245**, 123–126 (1998).
- Birse, D. E., Kapp, U., Strub, K., Cusack, S. & Åberg, A. The crystal structure of the signal recognition particle *Alu* RNA binding heterodimer, SRP9/14. *EMBO J.* **16**, 3757–3766 (1997).
- Weichenrieder, O., Kapp, U., Cusack, S. & Strub, K. Identification of a minimal *Alu* RNA folding domain that specifically binds SRP9/14. *RNA* **3**, 1262–1274 (1997).
- Strub, K., Moss, J. & Walter, P. Binding sites of the 9- and 14-kilodalton heterodimeric protein subunit of the signal recognition particle (SRP) are contained exclusively in the *Alu* domain of SRP RNA and contain a sequence motif that is conserved in evolution. *Mol. Cell. Biol.* **11**, 3949–3959 (1991).
- Cusack, S. *et al.* Small is beautiful: protein micro-crystallography. *Nature Struct. Biol.* **5**, 634–637 (1998).
- Pley, H. W., Flaherty, K. M. & McKay, D. B. Three-dimensional structure of a hammerhead ribozyme. *Nature* **372**, 68–74 (1994).
- Scott, W. G., Finch, J. T. & Klug, A. The crystal structure of an all-RNA hammerhead ribozyme: a proposed mechanism for RNA catalytic cleavage. *Cell* **81**, 991–1002 (1995).
- Wimberly, B. T., Guymon, R., McCutcheon, J. P., White, S. W. & Ramakrishnan, V. A detailed view of the ribosomal active site: the structure of the L11–RNA complex. *Cell* **97**, 491–502 (1999).
- Conn, G. L., Draper, D. E., Lattman, E. E. & Gittis, A. G. Crystal structure of a conserved ribosomal protein–RNA complex. *Science* **284**, 1171–1174 (1999).
- Strub, K. & Walter, P. Assembly of the *Alu* domain of the signal recognition particle (SRP): dimerization of the two protein components is required for efficient binding to SRP RNA. *Mol. Cell. Biol.* **10**, 777–784 (1990).
- Siegel, V. & Walter, P. Each of the activities of signal recognition particle (SRP) is contained within a distinct domain: analysis of biochemical mutants of SRP. *Cell* **52**, 39–49 (1988).
- Zwieb, C., Müller, F. & Larsen, N. Comparative analysis of tertiary structure elements in signal recognition particle RNA. *Fold. Des.* **1**, 315–324 (1996).
- Chang, D. Y., Newitt, J. A., Hsu, K., Bernstein, H. D. & Marais, R. J. A highly conserved nucleotide in the *Alu* domain of SRP RNA mediates translation arrest through high affinity binding to SRP9/14. *Nucleic Acids Res.* **25**, 1117–1122 (1997).
- Strub, K. & Walter, P. Isolation of a cDNA clone of the 14-kDa subunit of the signal recognition particle by cross-hybridization of differently primed polymerase chain reactions. *Proc. Natl Acad. Sci. USA* **86**, 9747–9751 (1989).
- Andrews, D. W., Walter, P. & Ottensmeyer, F. P. Evidence for an extended 7SL RNA structure in the signal recognition particle. *EMBO J.* **6**, 3471–3477 (1987).
- Nakamura, K., Yahagi, S., Yamazaki, T. & Yamane, K. *Bacillus subtilis* histone-like protein, HBSu, is an integral component of a SRP-like particle that can bind the *Alu* domain of small cytoplasmic RNA. *J. Biol. Chem.* **274**, 13569–13576 (1999).
- Strub, K., Fornallaz, M. & Bui, N. The *Alu* domain homolog of the yeast signal recognition particle consists of an Srp14p homodimer and a yeast-specific RNA structure. *RNA* **5**, 1333–1347 (1999).
- Siegel, V. & Walter, P. Binding sites of the 19-kDa and 68/72-kDa signal recognition particle (SRP) proteins on SRP RNA as determined in protein–RNA “footprinting”. *Proc. Natl Acad. Sci. USA* **85**, 1801–1805 (1988).
- Mighell, A. J., Markham, A. F. & Robinson, P. A. *Alu* sequences. *FEBS Lett.* **417**, 1–5 (1997).
- Leslie, A. G. W. *Joint CCP4 & ESRF-EACBM Newsletter on Protein Crystallography* No. 26 (Daresbury Laboratory, Warrington, UK, 1992).
- Navaza, J. AMoRe: an automated package for molecular replacement. *Acta Crystallogr. A* **50**, 157–163 (1994).
- Brünger, A. T. *et al.* Crystallography & NMR system: A new software suite for macromolecular structure determination. *Acta Crystallogr. D* **54**, 905–921 (1998).
- Perrakis, A., Morris, R. & Lamzin, V. S. Automated protein model building combined with iterative structure refinement. *Nature Struct. Biol.* **6**, 458–463 (1999).
- Jones, T. A., Zou, J. Y., Cowan, S. & Kjeldgaard, M. Improved methods for building protein models in electron density maps and the location of errors in these models. *Acta Crystallogr. A* **47**, 110–119 (1991).
- Otwinski, Z. & Minor, W. Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol.* **276**, 307–326 (1997).

43. de la Fortelle, E. & Bricogne, G. Maximum-likelihood heavy-atom parameter refinement for multiple isomorphous replacement and multiwavelength anomalous diffraction methods. *Methods Enzymol.* **276**, 472–494 (1997).
44. Abrahams, J. P. & Leslie, A. G. W. Methods used in the structure determination of bovine mitochondrial F1 ATPase. *Acta Crystallogr. D* **52**, 30–42 (1996).
45. Carson, M. RIBBONS 2.0. *J. Appl. Crystallogr.* **24**, 958–961 (1991).
46. Esnouf, R. M. An extensively modified version of MolScript that includes greatly enhanced coloring capabilities. *J. Mol. Graph.* **15**, 133–138 (1997).
47. Merritt, E. A. & Bacon, D. J. Raster3D: photorealistic molecular graphics. *Methods Enzymol.* **277**, 505–524 (1997).
48. Nicholls, A., Sharp, K. A. & Honig, B. Protein folding and association: insight from the interfacial and thermodynamic properties of hydrocarbon. *Proteins* **11**, 281–296 (1991).

Supplementary information is available on Nature's World-Wide Web site (<http://www.nature.com>) or as paper copy from the London editorial office of Nature.

Acknowledgements

O.W. received a pre-doctoral fellowship from the Boehringer-Ingelheim-Fonds. K.S. is supported by a START subsidy and a grant from the Swiss National Science Foundation and the Canton de Genève. K.S. and S.C. are both principal investigators belonging to SRPNET, a research network of the European Union Training and Mobility in Research program, which also supported K.W. We thank members of the EMBL/ESRF Joint Structural Biology Group (JSBG) and the staff of BM30 and ID13 for support on ESRF beamlines; in particular, A. Thompson for the europium MAD experiment and A. Perrakis for the ID13 measurements and help with wARP. C. Petosa commented critically on the manuscript.

Correspondence and requests for materials should be addressed to S.C. (e-mail: cusack@embl-grenoble.fr). Structures have been deposited in the PDB under accession numbers 1e8o (SA50 RNP) and 1e8s (SA88 RNP).

New high-pressure phases of lithium

M. Hanfland^{*}, K. Syassen[†], N. E. Christensen[‡] & D. L. Novikov[§]

^{*} European Synchrotron Radiation Facility, BP 220, 38043 Grenoble, France

[†] Max-Planck-Institut für Festkörperforschung, Heisenbergstr. 1, D-70569 Stuttgart, Germany

[‡] Institute of Physics and Astronomy, Aarhus University, DK-8000 Aarhus C, Denmark

[§] Arthur D. Little Inc., Acorn Park, Cambridge, Massachusetts 02140-2390, USA

Lithium is considered a 'simple' metal because, under ordinary conditions of pressure and temperature, the motion of conduction electrons is only weakly perturbed by interactions with the cubic lattice of atomic cores. It was recently predicted¹ that at pressures below 100 GPa, dense Li may undergo several structural transitions, possibly leading to a 'paired-atom' phase with low symmetry and near-insulating properties. Here we report synchrotron X-ray diffraction measurements that confirm that Li undergoes pronounced structural changes under pressure. Near 39 GPa, the element transforms from a high-pressure face-centred-cubic phase, through an intermediate rhombohedral modification, to a cubic polymorph with 16 atoms per unit cell. This cubic phase has not been observed previously in any element; unusually, its calculated electronic density of states exhibits a pronounced semimetal-like minimum near the Fermi energy. We present total-energy calculations that provide theoretical support

for the observed phase transition sequence. Our calculations indicate a large stability range of the 16-atom cubic phase relative to various other crystal structures tested here.

The predictions by Neaton and Ashcroft¹ are in sharp contrast to intuitive expectation that the application of hydrostatic pressure favours high-coordination crystal structures with metallic properties. In their theoretical simulations of dense Li, which are based on first principles band structure theory, they compare the relative stability of a number of crystal structures common among elemental solids. Their results clearly indicate a strong preference of dense Li to form low-symmetry structures. Therefore, experiments aimed at structure determinations of compressed Li are highly desirable. Furthermore, experimental high-pressure studies of Li are of fundamental interest, because they are expected to reveal new aspects relevant for the theoretical modelling of other light elements, including hydrogen², at high density.

We have performed monochromatic synchrotron X-ray powder diffraction studies of Li in a diamond anvil cell (DAC) at pressures up to 50 GPa which is significantly higher compared to previous diffraction studies of Li (refs 3–5). The procedures for data collection and analysis were largely similar to those employed in recent structural studies of heavy alkali metals^{6,7}. Lithium (⁷Li, stated purity 99.9%) was loaded into a DAC under inert gas. No pressure medium was used in order to avoid chemical reactions. A ruby chip served as *in situ* optical pressure sensor. Limitations arising from the unusual reactivity of Li under pressure, causing for instance the failure of diamond anvils⁴, were overcome by performing the experiments at reduced temperatures. The DAC was placed in a flow cryostat and, for all pressures above 20 GPa, was kept at a temperature not exceeding 200 K. Data were collected for several sample loadings, yielding fully reproducible results.

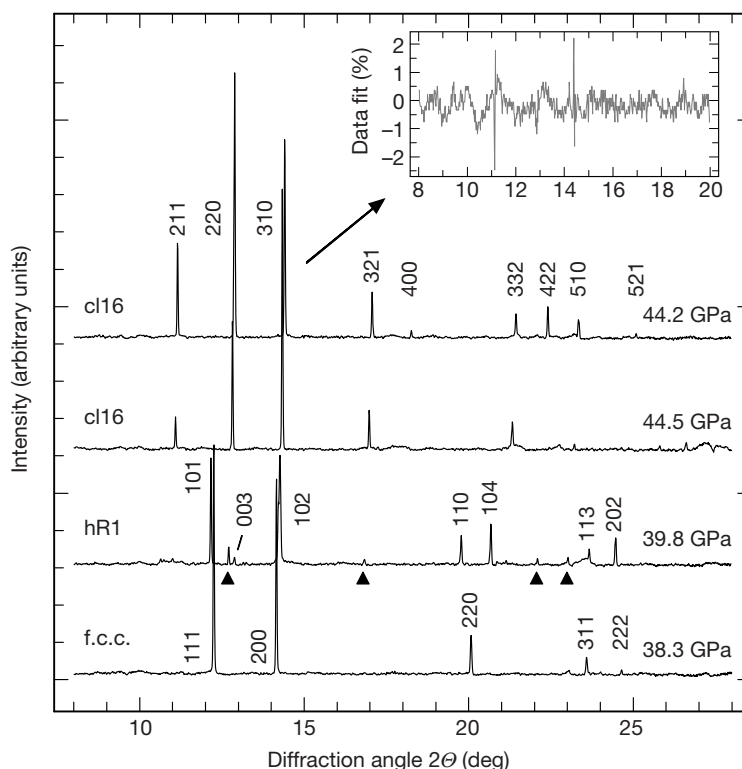


Figure 1 Synchrotron X-ray diffraction diagrams of high-pressure phases of lithium. The angle-dispersive powder diffraction experiments (wavelength 0.4124 Å) were carried out at the undulator beamline ID9 of the European Synchrotron Radiation Facility using a diamond anvil high-pressure cell at 180 K. Diffraction patterns were recorded on a flat image plate and then converted to the intensity versus 2θ diffraction angle data shown here. A smooth background, arising mainly from Compton scattering in the diamond

windows, has been removed. The sequence of diagrams is for the f.c.c. phase (38.3 GPa), the Li-hR1 phase (39.8 GPa) with a small admixture of Li-cl16 (marked by triangles), and the Li-cl16 phase (42.5 GPa and 44.2 GPa). The inset shows the difference, in percentage of the maximum peak intensity, between data and refined profile for the diagram at 44.2 GPa.

In addition to the body-centred cubic (b.c.c.) phase, two low-temperature modifications of Li are known to occur near ambient pressure, the Sm-type (9R) and face-centred-cubic (f.c.c.) phases^{8–10}. High-pressure X-ray diffraction experiments at 300 K have revealed a b.c.c. $\rightarrow$ f.c.c. transition near 7.5 GPa (refs 3–5). Representative diffraction data obtained in our experiments at 180 K are shown in Fig. 1. The structure remained f.c.c. up to 38.3 GPa. At 39.8 GPa a new set of intense diffraction lines was observed together with a second set of weak-intensity lines. The first new set disappeared at slightly higher pressure (42.5 GPa), the other one gained intensity. We conclude that Li transforms to an intermediate phase between 39 and 42 GPa and another new phase emerges in this pressure range which is stable up to at least 50 GPa.

The Bragg reflections of the intermediate phase can be indexed in the hexagonal system with lattice parameters $a = 2.4023 \text{ \AA}$, $c = 5.51592 \text{ \AA}$, and $c/a = 2.296$ (at 39.8 GPa). The c/a ratio is strongly pressure dependent (2.250 at 42.5 GPa). Comparison with the extrapolated atomic volume of the f.c.c. phase suggests three atoms in the hexagonal cell. The systematic extinctions are

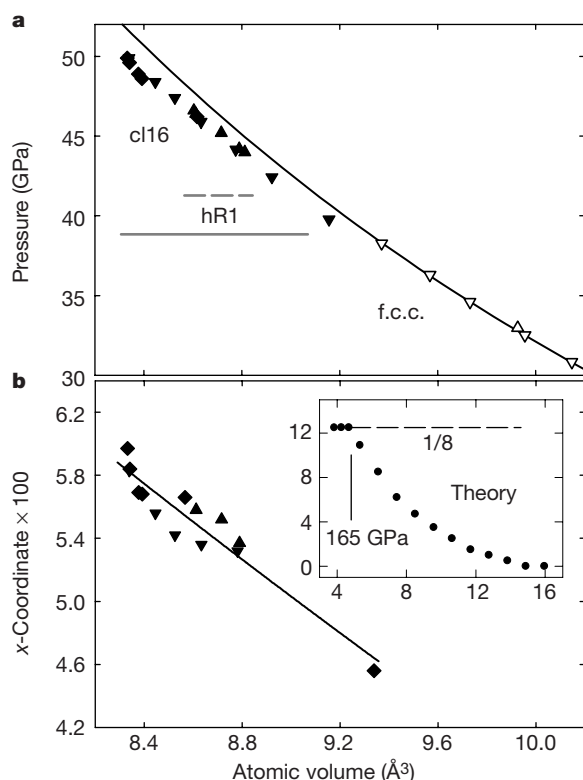


Figure 2 Structural parameters of compressed lithium. Experimental pressure–volume data for the f.c.c. (open symbols) and cl16 (closed symbols) phases in **a** are from different experiments performed at temperatures T between 160 K and 200 K. The solid line refers to a semi-empirical pressure–volume relation fitted to data for the f.c.c. phase in the range 10–39 GPa. Horizontal lines indicate the pressure range for which the hR1 phase coexists with the cl16 phase for increasing pressure. At 39.8 GPa, where Li is compressed to about 42% of its ambient pressure volume, the atomic volumes are 9.2395, 9.1893 and 9.1560 $\text{\AA}^3$ for the f.c.c., hR1 and cl16 phases, respectively. Thus, near 40 GPa the atomic volumes of the three phases differ by less than 1%. **b**, The refined positional parameter x for the Li-cl16 phase as a function of atomic volume. The outlier data point representing the smallest observed x -value refers to an experiment performed at $T = 100 \text{ K}$ and in the direction of decreasing pressure. At this temperature the diffraction pattern of the cl16 phase was still fully developed at 37.5 GPa, that is, at a lower pressure compared to the first appearance of the cl16 phase for increasing pressure. It remains to be investigated if this indicates hysteresis effects or a positive slope of the cl16 stability line in the P – T phase diagram. The inset to **b** shows the calculated variation of the structural parameter x of the cl16 structure over a large volume range corresponding to pressures up to 200 GPa.

consistent with space group (SG) $R\bar{3}m$. The corresponding rhombohedral cell contains one atom (Pearson symbol¹¹ hR1). The structure may be regarded as a distortion of f.c.c. involving a compression of about 7% along one of the three-fold cubic axes and an increase of interatomic distances in hexagonal layers perpendicular to that axis. Mercury, for example, crystallizes in a similar structure ($c/a = 2.185$).

The second new phase of Li, however, adopts a new structure type, so far not found for any other element. All the diffraction lines observed for this phase index for a cubic cell ($a = 5.2716 \text{ \AA}$ at 38.9 GPa). The highest-symmetry space group consistent with the systematic extinctions is $I\bar{4}3d$. On the basis of volume considerations, a Wyckoff site 16c, having fractional coordinates (x, x, x), is the only possible choice. Thus, the new cubic structure of Li has a body-centred unit cell with 16 atoms (Pearson symbol cI16). Selected pressure versus atomic volume data for this phase are shown in Fig. 2a. Diffraction patterns of Li-cl16 were refined by the Rietveld method, allowing for preferred orientation effects. The obtained values of the internal positional parameter x vary between 0.045 and 0.060, showing a clear tendency to increase with decreasing volume (Fig. 2b).

The crystal structure of Li-cl16 (Fig. 3) is easily derived from a simple b.c.c. lattice (SG $Ia\bar{3}d$). The latter can be regarded as a packing of non-intersecting cylinders (chains of equidistant atoms) running parallel to the body diagonals (the four equivalent $\langle 111 \rangle$ directions) of the cubic cell^{12,13}. A shift of rigid cylinders with respect to each other lowers the symmetry to SG $I\bar{4}3d$ and results in a $2 \times 2 \times 2$ supercell formation of b.c.c. The coordination changes from $8 + 6$ for b.c.c. to $3 + 2 + 6$ for Li-cl16 with $x = 0.05$. The structure of a high pressure form of Ga belongs to the same space group, but Ga atoms occupy the 12a site. The arrangement of atoms in Li-cl16 is very similar to that of cation sublattices in compounds with the anti- Th_3P_4 -type and Pu_2C_3 -type structures¹³, for example, Yb_4As_3 or Rb_4O_6 . Thus, a trend emerging from structural studies of the heavy alkali metals also applies to Li; the low-

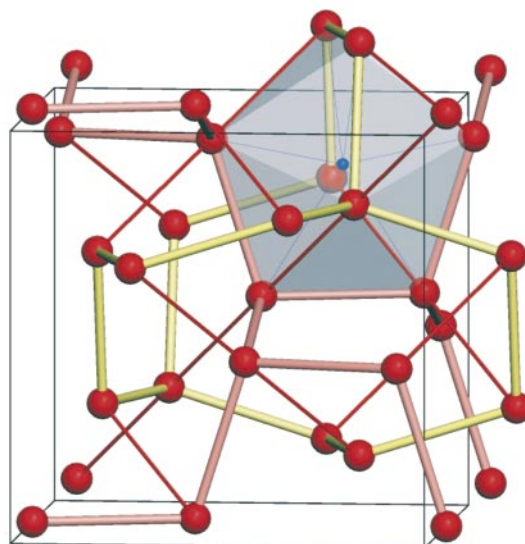


Figure 3 Schematic representation of the cubic crystal structure of Li near 45 GPa. The space group is $I\bar{4}3d$. The cubic cell contains 16 atoms occupying the Wyckoff 16c site (positional parameter $x = 0.055$). The shortest interatomic distances (2.1 $\text{\AA}$) are indicated by thick lines. The thinner ones connect equidistant atoms on chains running parallel to the four differently oriented body diagonals of the cubic lattice. The structure can be viewed as two interpenetrating 3-connected nets, as indicated by different colours for the thick lines. An alternative interpretation is in terms of a framework of distorted Li_8 dodecahedra. One of the Li_8 polyhedra is indicated by shading. For $x = 0.0833$ ($1/12$) the eight atoms at the vertices of a polyhedron would all have the same distance from the centre position, which corresponds to a Wyckoff 12a site.

symmetry structures encountered when these 'simple metals' are compressed^{7,14–16} have counterparts in cation sublattices of binary compounds^{7,17}.

In our theoretical modelling of the phase stability of Li the total electron energy E for a given crystal structure is calculated by density functional theory using the generalized gradient approach¹⁸. The solution of the effective one-electron equations is performed by means of the linear muffin-tin orbital method¹⁹ in the full potential version (M. Van Schilfgaarde and M. Methfessel, unpublished work). In the scalar-relativistic all-electron calculations the $1s$ states were treated as local orbitals²⁰ forming bands. A "triple- κ " basis²¹ with s -, p -, d - and f -states in each channel was used. The muffin-tin sphere radius R_{mt} was chosen to be as large as possible, that is, adjusted to the smallest interatomic separation occurring within the tested range of structural parameters for a given structure. R_{mt} scales with volume so that the sphere volume is a constant fraction of the total volume. Structural energy differences between a given structure and the f.c.c. structure (used as reference) are calculated by comparing results obtained with the same sphere

volume packing ratio in both structures. At a given volume V the unit cell parameters and atom positions (if applicable) were optimized simultaneously by a steepest descent method. Having calculated the optimized E versus V relation for a given modification, we derive the pressure–volume (P – V) relation which is then used to calculate the enthalpy $H(P) = E(P) + PV$.

The calculated real-space electron distribution for the optimized Li-cl16 structure at $V/V_0 = 0.4$ (Fig. 4a) shows pronounced maxima in the interstitial regions. In chemical terms the charge density indicates a multi-centre bonding situation. The unusual charge distribution in Li-cl16 is reflected in the electronic structure (Fig. 4b). The calculated electronic density of states (DOS) for the undistorted b.c.c. superlattice ($x = 0.0$) peaks close to the Fermi level E_F which is due to a flat degenerate band near E_F . The distortion towards $x = 0.05$ lifts the band degeneracy and changes the band dispersions, causing a pronounced minimum in the DOS close to E_F and a maximum about 0.4 eV below E_F .

One would expect major changes in the electrical transport and optical properties of Li near the transition to the cl16 phase. The

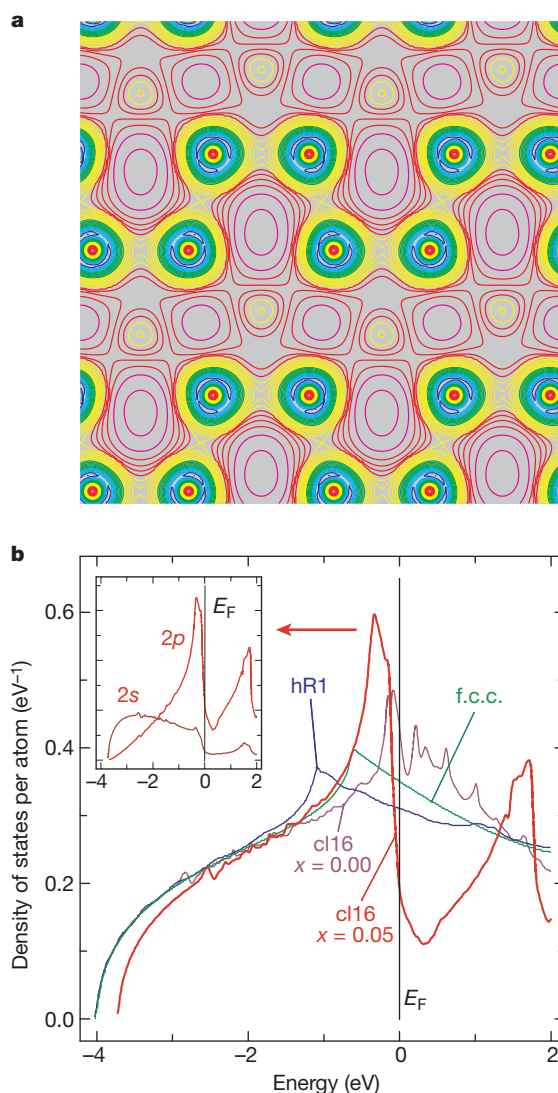


Figure 4 Calculated charge density distribution and electronic density of states of Li at a relative volume $V/V_0 = 0.4$ (theoretical pressure 48.8 GPa). The charge density plot in **a** covers four unit cells in a (100) plane passing through the origin at (0,0,0) of the cl16 structure. Contributions of $1s$ states are subtracted. The colour coding is blue towards red/magenta for increasing electron density. The density is higher than the average value ($\rho_{\text{av}} = 0.1157 \text{ e}\text{\AA}^{-3}$) in all areas marked by red/magenta contours. The maximum

density in the interstitial regions is $1.50\rho_{\text{av}}$. The results in **b** demonstrate that the $x = 0$ to 0.05 distortion of the cl16 structure leads to a pronounced maximum in the density of states just below the Fermi level E_F followed by a minimum near E_F . In this way the total electronic energy is lowered with respect to a b.c.c. superlattice ($x = 0$) and also with respect to the f.c.c. and hR1 structures. The inset to **b** shows relative contributions of s - and p -like orbitals to the density of states.

sharp increase in the electrical resistance²² observed for shock-compressed Li at densities of 1.2–1.4 g cm⁻³ appears to be related to the structural behaviour reported here. Within the rather narrow spectral window of visual observations Li remained highly reflecting up to 50 GPa, showing, however, some changes in its metallic lustre at the phase transitions. In this context we note that neither at room temperature nor below have we been able to reproduce the change to black colour at 20–25 GPa as observed in two recent experimental studies^{5,23} of Li, both covering the pressure range up to 120 GPa.

With respect to b.c.c., the distortion towards Li-cl16 is reminiscent of a Peierls-type instability where a symmetry-breaking structural distortion leads to a gain in band structure energy. The accumulation of electron density in the interstitial regions of Li-cl16 could also be viewed as a commensurate charge density wave (CDW) formed in a (hypothetical) b.c.c. lattice, in particular because the calculated volume difference between Li-b.c.c. and Li-cl16 is also very small. Overhauser has considered incommensurate CDW instabilities in alkali metals at normal density, mainly driven by exchange interactions²⁴. Related to this, the redistribution of electron density at the transition to Li-cl16 possibly results in a gain of nonlocal exchange energy, which is only approximately included in our theoretical model¹⁸.

Boettger and Trickey²⁵ were probably first to show that the calculated band structure of compressed Li (f.c.c. assumed) deviates significantly from the nearly free electron behaviour. They inter-

preted their results in terms of a pressure-driven transition from *s*-towards *p*-like orbital character of the occupied states. This effect also applies to Li-cl16; the *p*-orbital component is particularly large for the pronounced DOS maximum just below *E*_F (see inset to Fig. 4b). Conceptually, the change in the conduction electron orbital character in Li is similar to the pressure-driven *s* → *d* transition in heavy alkali metals^{26,27}. Their low-pressure f.c.c. modifications transform to low-coordination phases^{7,14–16}, similar to the case of Li.

Optimized structural parameters for Li-cl16 were calculated for an extended range of volumes. In agreement with experiment the *x*-parameter increases with decreasing volume (see inset to Fig. 2b). At a calculated pressure of 165 GPa (*V/V*₀ = 0.23) it reaches a value of 1/8, which actually corresponds to a structure with SG *Ia-3d* (atoms in 16b). In this structural arrangement Li remains a semimetal, but the DOS at *E*_F is significantly reduced compared to that at *V/V*₀ = 0.4 (Fig. 4b). The nearest neighbour coordination consists of three coplanar atoms in planes perpendicular to body diagonals and two out-of-plane atoms at slightly larger distance. In other words, with increasing pressure the optimized Li-cl16 structure may change continuously from eleven-fold coordination near 40 GPa towards five-fold coordination at 165 GPa. This leads to the question of whether the cl16 phase of Li would be stable up to this pressure.

We have calculated differences in enthalpy, relative to f.c.c., for several test structures and covering the range up to 200 GPa (Fig. 5). Taking into account numerical errors of 5 meV per atom, the calculated phase transition sequence near 40–50 GPa (see inset to Fig. 5) is in good agreement with the experimental results. With respect to all other structures considered here the optimized Li-cl16 structure is lowest in enthalpy from 48 GPa up to 165 GPa, where it yields to the oC8 (Cmca) ‘paired-atom’ phase predicted in ref. 1. However, the transition to Li-oC8 is shifted up in pressure by roughly 50 GPa compared to a possible A7 (arsenic-type) → oC8 transition envisioned in ref. 1. In agreement with ref. 1 we find strongly compressed Li-oC8 (*V/V*₀ = 0.25) to be a near-zero-gap semiconductor. The optimized Li-oC8 structure has one short interatomic distance of 1.47 Å and four more short contacts at 1.69 to 1.79 Å. Thus, the coordination again is close to five-fold as in the optimized Li-cl16 phase at 165 GPa. The structure of Li-oC8 resembles that of α-gallium, but can also be viewed as that of black phosphorus compressed in the direction perpendicular to its double layers.

The structural sequence observed for Li up to 50 GPa, while following a different route from that predicted by Neaton and Ashcroft¹, nevertheless represents an experimental confirmation of their general conclusions that dense Li becomes unstable towards symmetry-breaking transitions at pressures near 50 GPa. Further experiments at lower temperatures will help to clarify the lattice dynamical effects in the observed phase transitions. A particularly attractive subject would be to extend the search for superconductivity in Li under pressure²⁸ into the phase transition regime near 40 GPa. In our calculations the Li-cl16 phase shows a very large stability region, yielding to a oC8 (Cmca) phase¹ only at a pressure near 165 GPa. Depending on temperature and/or pressure, new phases may interfere and the structural behaviour of dense Li will remain a subject for future experimental and theoretical investigations of this unusual metal. □

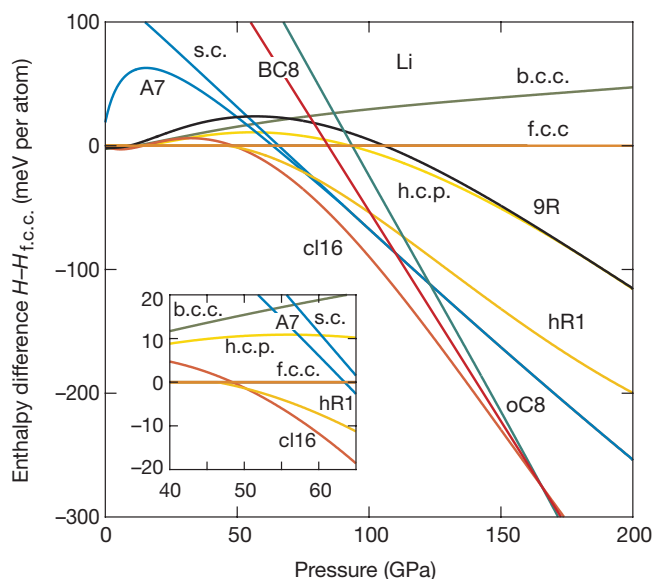


Figure 5 Calculated enthalpy differences (relative to f.c.c.) for Li in various crystal structures as a function of pressure. These are b.c.c., h.c.p., 9R (Sm-type), hR1, Li-cl16, s.c. (simple cubic), A7 (arsenic-type) and oC8 (Cmca) (see text). Lattice dynamical effects (zero-point and thermal motion of atoms) are not included in these calculations. Where relevant, all structural parameters were optimized. Thus, relaxation of the A7 towards the s.c. structure explains the confluence of the corresponding curves above 80 GPa. The predicted phase transition sequence with increasing pressure is 9R → 4 GPa → b.c.c. → 16 GPa → f.c.c. → 41 GPa → hR1 → 48 GPa → cl16 (I-43d) → 165 GPa → cl16 relaxed towards SG *Ia-3d* → >165 GPa → oC8. Two structures closely related to that of Li-cl16, the cubic BC8 (SG *Ia-3*) and rhombohedral R8 (SG *R-3*) structures of metastable Si (ref. 29), were also considered. Throughout the density range investigated here the BC8 and R8 structures of Li are found to relax towards SG *Ia-3d* and their enthalpy (curve labelled BC8) is higher than that of Li-cl16, except near 165 GPa where the optimized Li-cl16 also approaches SG *Ia-3d*. Evaluation of the enthalpy results and transition pressures should take into account that our calculated enthalpy differences have estimated error bars of the order of 5 meV per atom. At zero pressure the 9R–f.c.c. and b.c.c.–f.c.c. energy differences are 1 to 5 meV per atom, and on that energy scale we found that the choice of band structure method and exchange–correlation potential can change the ordering. This does not affect our conclusions about the high-pressure structures.

Received 22 May; accepted 20 September 2000.

- Neaton, J. B. & Ashcroft, N. W. Pairing in dense lithium. *Nature* **400**, 141 (1999).
- Martin, R. M. Simple metals under pressure. *Nature* **400**, 117–119 (1999).
- Olinger, B. & Shaner, W. Lithium, compression and high-pressure structure. *Science* **219**, 1071 (1983).
- Hanfland, M., Loa, I., Syassen, K., Schwarz, U. & Takemura, K. Equation of state of lithium to 21 GPa. *Solid State Comm.* **112**, 123–127 (1999).
- Struzhkin, V. V., Hemley, R. J. & Mao, H. K. Compression of lithium to 120 GPa. *Bull. Am. Phys. Soc.* **44**, 1489 (1999).
- Schwarz, U., Takemura, K., Hanfland, M. & Syassen, K. Crystal structure of cesium-V. *Phys. Rev. Lett.* **81**, 2711–2714 (1998).
- Schwarz, U., Grzechnik, A., Syassen, K., Loa, I. & Hanfland, M. Rubidium-IV: a high pressure phase

- with complex crystal structure. *Phys. Rev. Lett.* **83**, 4085–4088 (1999).
8. Overhauser, A. W. Crystal structure of lithium at 4.2 K. *Phys. Rev. Lett.* **53**, 64–65 (1984).
 9. Vaks, V. G. *et al.* An experimental and theoretical study of martensitic phase transitions in Li and Na under pressure. *J. Phys. Condens. Matter* **1**, 5319–5335 (1989).
 10. Smith, H. G., Berliner, R., Jorgensen, J. D., Nielsen, M. & Trivisonno, J. Pressure effects on the martensitic transformation in metallic lithium. *Phys. Rev. B* **41**, 1231–1234 (1990).
 11. Pearson, W. B. *A Handbook of Lattice Spacings and Structures of Metals and Alloys* Vol. 2 (Pergamon, Oxford, 1967).
 12. O'Keeffe, M. & Hyde, B. G. *Crystal Structures* (Mineralogical Society of America, Washington DC, 1996).
 13. Wells, A. F. *Structural Inorganic Chemistry* 5th edn (Oxford Univ. Press, Oxford, 1984).
 14. Takemura, K., Minomura, S. & Shimomura, O. X-ray diffraction study of electronic transitions in cesium under high pressure. *Phys. Rev. Lett.* **49**, 1772–1775 (1982).
 15. Olijnyk, H. & Holzapfel, W. B. Phase transitions in K and Rb under pressure. *Phys. Lett. A* **99**, 381–386 (1983).
 16. Winzenick, M., Vijayakumar, V. & Holzapfel, W. B. High pressure x-ray diffraction of potassium and rubidium up to 50 GPa. *Phys. Rev. B* **50**, 12381–12385 (1994).
 17. Von Schnering, H. G. & Nesper, R. How nature adapts chemical structures to curved surfaces. *Angew. Chem. Int. Edn Engl.* **26**, 1059–1080 (1987).
 18. Perdew, J. P., Burke, K. & Ernzerhof, M. Generalized gradient approximation made simple. *Phys. Rev. Lett.* **77**, 3865–3868 (1996).
 19. Andersen, O. K. Linear methods in band theory. *Phys. Rev. B* **12**, 3060–3083 (1975).
 20. Singh, D. J. *Planewaves, Pseudopotentials and the LAPW Method* (Kluwer, Boston, 1994).
 21. Methfessel, M. Elastic constants and phonon frequencies of Si calculated by a fast full-potential linear-muffin-tin-orbital method. *Phys. Rev. B* **38**, 1537–1540 (1988).
 22. Fortov, V. E. *et al.* Anomalous electrical conductivity of lithium under quasi-isotropic compression to 60 GPa (0.6 Mbar). Transition into a molecular phase? *JETP Lett.* **70**, 628–632 (1999).
 23. Mori, Y., Zha, C.-S. & Ruoff, A. L. in *Science and Technology of High Pressure* (eds Manghani, M. H., Nellis, W. J. & Nicol, M. F.) (Univ. Press India, Hyderabad, 2000).
 24. Overhauser, A. W. Exchange and correlation instabilities of simple metals. *Phys. Rev.* **167**, 691–698 (1968).
 25. Boettger, J. C. & Trickey, S. B. Equation of state and properties of lithium. *Phys. Rev. B* **32**, 3391–3398 (1985).
 26. Sternheimer, R. On the compressibility of metallic cesium. *Phys. Rev.* **78**, 235–243 (1950).
 27. McMahon, A. K. Alkali metal structures above the s-d transition. *Phys. Rev. B* **29**, 5982–5985 (1984).
 28. Lin, T. H. & Dunn, K. J. High-pressure and low-temperature study of the electrical resistance of lithium. *Phys. Rev. B* **33**, 807–811 (1986).
 29. Nemes, R. J. & McMahon, M. I. in *High Pressure in Semiconductor Physics* Vol. 1, 146–246 (Academic, New York, 1998).

Correspondence should be addressed to K.S. (e-mail: syassen@servix.mpi-stuttgart.de).

Reversible phase transitions in polymer gels induced by radiation forces

Saulius Juodkazis*, Naoki Mukai†, Ryosuke Wakaki†, Akira Yamaguchi*, Shigeki Matsuo† & Hiroaki Misawa†

* Satellite Venture Business Laboratory of Photonic Nano-Materials, The University of Tokushima, 2-1 Minamijosanjima, Tokushima 770-8506, Japan

† Department of Ecosystem Engineering, Graduate School of Engineering, The University of Tokushima, 2-1 Minamijosanjima, Tokushima 770-8506, Japan

Many polymer gels undergo reversible, discontinuous volume changes in response to changes in the balance between repulsive intermolecular forces that act to expand the polymer network and attractive forces that act to shrink it. Repulsive forces are usually electrostatic or hydrophobic in nature, whereas attraction is mediated by hydrogen bonding or van der Waals interactions. The competition between these counteracting forces, and hence the gel volume^{1–3}, can thus be controlled by subtle changes in parameters such as pH (ref. 4), temperature⁵, solvent composition⁶ or gel composition⁷. Here we describe a more direct influence on this balance of forces, by showing that the radiation force generated by a focused laser beam induces reversible shrinkage in polymer gels. Control experiments confirm that the laser-induced volume phase transitions are due to radiation

forces, rather than local heating, modifying the weak interactions in the gels, in agreement with previous observations of light-induced chain association in polymer solutions^{8,9}. We find that, owing to shear-relaxation processes¹⁰, gel shrinkage occurs up to several tens of micrometres away from the irradiation spot, raising the prospect that the combination of stimuli-responsive polymer gels and laser light might lead to new gel-based systems for applications such as actuating or sensing.

Laser trapping is a valuable tool for basic and applied research^{11–13}. It uses radiation force to immobilize particles against brownian motion and any convection¹⁴. When a laser beam of wavelength λ is focused on a particle of radius r ($r \ll \lambda$), it exerts a radiation force that can be expressed in terms of the gradient and scattering forces as, F_G and F_P , respectively, as $F = F_G + F_P$, given by¹⁵

$$F = n\alpha \left(\frac{\nabla E^2}{2} + \frac{\partial}{\partial t} (\mathbf{E} \times \mathbf{B}) \right) \quad (1)$$

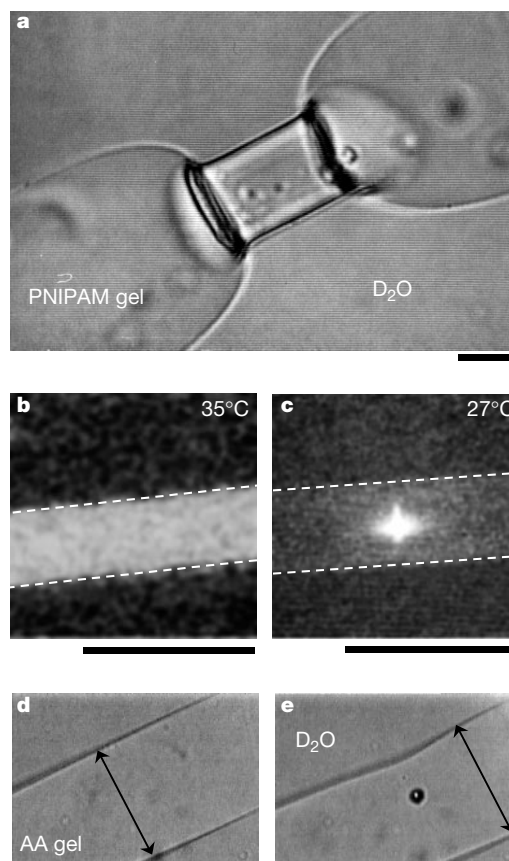


Figure 1 The volume phase transition in gels. **a**, The volume phase transition of a laser-illuminated rod of PNIPAM gel in D₂O. The fluorescent probe 8-anilino-1-naphthalene-sulphonic acid sodium salt (ANS-Na), which is sensitive to the hydrophobicity of a microenvironment, was chosen to demonstrate light-induced compaction of gels. We measured its photoluminescence (PL) at its wavelength of maximum intensity (485 nm; PL was excited by Hg-lamp irradiation at 365 nm). The PNIPAM gel was soaked in a D₂O solution of the dye. The walls of the capillary were chemically modified by a silane-coupler, which created a hydrophobic surface that prevented the gel from collapsing under the conditions of a volume phase transition (for example, at a temperature above T_{tr} : $T_{tr} \approx 34^\circ\text{C}$). **b**, The PL of ANS-Na observed at temperatures above T_{tr} . This corroborates a compaction of the gel on a submicrometre scale due to segment–segment interaction of the polymer coils inside the gel. The PL was observed all along the capillary. It can be seen that the dye's PL was also enhanced at temperatures lower than T_{tr} (**c**) in the vicinity of the focal point, where the radiation force induces the gel compaction. Such a compaction is demonstrated in acrylamide (AA) gel (at state laser ON (**d**) and OFF (**e**)) at the same illumination conditions (D₂O immersion; similar absorption coefficients of AA and PNIPAM at 1,064 nm wavelength; the same 1.2 W illumination power). Scale bars, 20 μm .

where the polarizability of the particle under the dipole approximation is $\alpha = [(n_p/n)^2 - 1]/[(n_p/n)^2 + 2]$. Here n and n_p are refractive indices of the surrounding medium and particle, respectively. $\mathbf{E}$ and $\mathbf{B}$ are electric field strength and magnetic flux density, respectively. When a non-absorbing particle with $n_p > n$ is illuminated by the tightly focused laser light, a force gradient (the first term in the brackets of equation (1)) usually dominates, and the particle is three-dimensionally captured at the locus of the region of high light intensity.

We prepared rod-shaped microgels by free-radical polymerization⁷ (at 27 °C for 1 h) of a solution of *N*-isopropylacrylamide (NIPAM) in D₂O; the solution was inside a glass microcapillary tube. The resulting poly(*N*-isopropylacrylamide) (PNIPAM) gel was taken out of the tube, then soaked and washed in D₂O for 3 d. We found that the radiation force of a single-beam laser trap could trigger gel compaction (a volume phase transition) in PNIPAM (Fig. 1a); we observed a similar effect in a PNIPAM-SA copolymer gel (here SA indicates sodium acrylate). It was necessary to immerse the free gels in D₂O, and to use D₂O throughout rather than H₂O, to avoid heating caused by absorption of the 1,064-nm laser light by an overtone of the H–O vibration band. Such an absorption alone can cause the temperature to rise, as we found earlier in an aqueous PNIPAM solution¹⁶ where the heating was responsible for the aggregation of polymer chains in the vicinity of the focus.

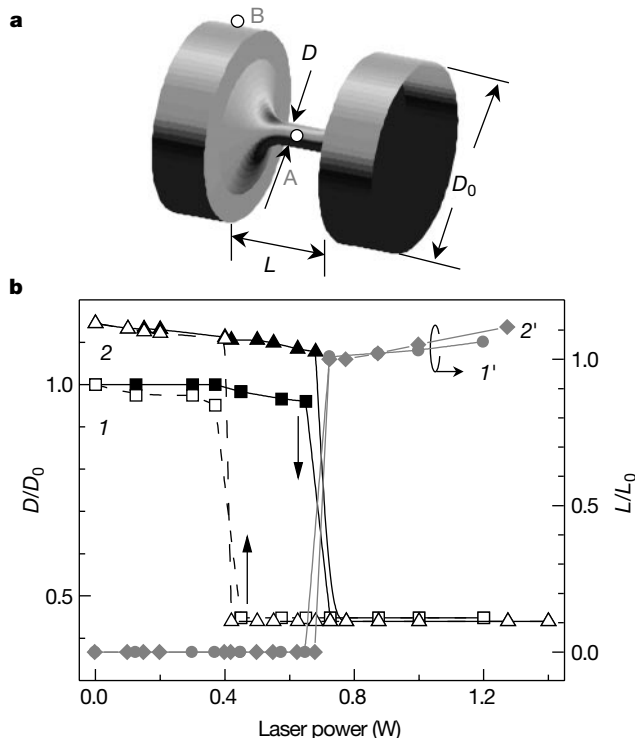


Figure 2 The power dependence of gel-rod collapse. **a**, The gel rod with its partly completed volume phase transition, which was induced at point A by radiation force. No volume phase transition was induced when point B was illuminated to check the direct D₂O heating by laser light absorption. **b**, The dependence of the diameter D/D_0 (left-hand vertical axis) and the waist length of the collapsed part L/L_0 (right-hand axis) of PNIPAM gel versus the power of the laser irradiation at 1,064 nm in D₂O at 27 °C. Curves 1 and 2 correspond to the pure PNIPAM gel and that doped by 4.6 wt% of sodium acrylate (SA), respectively. Solid lines and filled symbols correspond to the increasing-power part of a cycle; the dashed line and the open symbols to the decreasing-power part. Repeated cycling did not change the thresholds of shrinkage and expansion, corroborating the temperature stability at the focal point. D_0 is the initial diameter of the gel rod, and L_0 is the length of the waist (the length of the collapsed gel) at time $t = 2.5$ s after the start of illumination (at this time the waist is fully formed).

The observed gel collapse is caused by segment–segment interactions within the gel, similar to those giving rise to coil-to-globule transitions in polymer solutions¹⁷. That is, the water molecules form hydrogen-bonded networks, which surround the hydrophobic isopropyl segments of PNIPAM and thus favour extended segment conformations; with increasing temperature, thermal fluctuations render the water network less stable, so that attractive interactions between hydrophobic segments eventually cause the coil to collapse into a globule. This transition is further facilitated by van der Waals forces between identical monomers being stronger than those between unlike pairs. Finally, these weak interactions cause volume collapse of PNIPAM gel at the specific volume phase transition temperature, T_{tr} .

To estimate the temperature at the focal point, we measured the intensity ratio of the photoluminescence (PL) of excited monomers at 415 nm to that of excimers at 485 nm in a D₂O solution of the dye 1-pyrenesulphonic acid sodium salt (PS-Na). A linear dependence on temperature was confirmed (see Supplementary Information), allowing us to infer from measured intensities the temperature inside the gel at the illuminated spot, and thus calculate the local temperature increase¹⁸ due to light absorption at 1,064 nm by D₂O and PNIPAM. We find this increase to be less than (1.7 ± 0.2) K for an irradiation power of 1.2 W, while the typical power employed to induce the volume-phase transition of the gel was 0.7–1 W. The absorption coefficient of D₂O at 1,000 nm wavelength is 0.1 cm^{-1} and that of PNIPAM gel is negligible.

The local PNIPAM gel compaction (the compaction of the polymer coils) is demonstrated by enhanced dye luminescence under the action of a temperature increase (Fig. 1b) or laser trap at the focal point (Fig. 1c). We found that the radiation force of the laser trap can induce the volume phase transition (Fig. 2) and, also, lower the transition temperature T_{tr} by about 10 °C at an irradiation power of 1.2 W in PNIPAM gel (Fig. 3). In the case of PNIPAM-SA gel, T_{tr} can be lowered by radiation force as well. A discontinuous volume phase transition can be observed at SA concentrations at which T_{tr} increases towards the boiling point of water, and usual heating (curve 3 in Fig. 3) cannot trigger the transition⁷. Illumination of the gel by the tightly focused laser beam at 20 °C created a small (1–3%) compaction, as can be seen in Fig. 3. The compaction was also found in acrylamide gels, in which the volume phase transition is not temperature sensitive (Fig. 1d, e).

The pressure exerted by radiation force is directional, and acts towards the centre of the illumination spot. Such a compaction of NIPAM gel (curve 1 in Fig. 3) can be expressed in terms of an

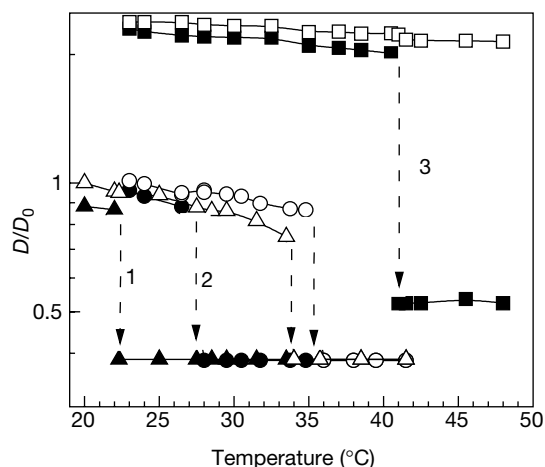


Figure 3 Temperature dependence of the diameter of rods of doped and undoped PNIPAM gel. Curves 1–3 correspond to sodium acrylate (SA) doping concentrations of 0, 4.6 and 18.3 wt%, respectively. Open and filled symbols depict measurements respectively without and with laser irradiation. Laser power, 0.75 W.

osmotic pressure of about 45 MPa, using the formulae developed in ref. 7 (Fig. 4). This is equivalent to the compaction of NIPAM gel that can be induced by a temperature increase of about 10 °C: that is, the osmotic pressure due to radiation force can drive a gel towards the volume phase transition, as does a temperature increase. The pressure due to the force gradient of the laser trap can be calculated¹¹; we obtain 4–250 MPa for spheres of 10–40 nm radius. (We utilized the following parameters in calculations using equation (1): light intensity, 1 W; refractive indexes of heavy water and NIPAM, respectively 1.326 and 1.37.) This corroborates our interpretation of a volume phase transition modified by radiation pressure, which lowers the transition temperature.

We also subjected polymer-gel rods to prolonged illumination (several minutes) while monitoring the system by video with a time resolution of 1/16 s. In these experiments, we used a laser power smaller than that required ($P_{cr} = 0.7$ W) to trigger a volume phase transition immediately. No waist formation—that is, no volume phase transition—was detected at $P < P_{cr}$. Once P_{cr} was reached (curves 1' and 2' in Fig. 2), gel collapse occurred by diffusional phase separation with a time dependence of $D/D_0 \propto t^{0.5}$, with D and D_0 being the diameter of laser illuminated and un-illuminated gel rod, respectively. Collapse typically proceeded within less than 2 s for the 20–30 μm diameter gel rods (see Supplementary Information), and resulted in the formation of a waist of length L_0 , with L_0 being independent of the illumination power up to 1 W. These data, together with the PL measurements of local heating, exclude temperature as the determining factor leading to the volume phase transition in our experiments.

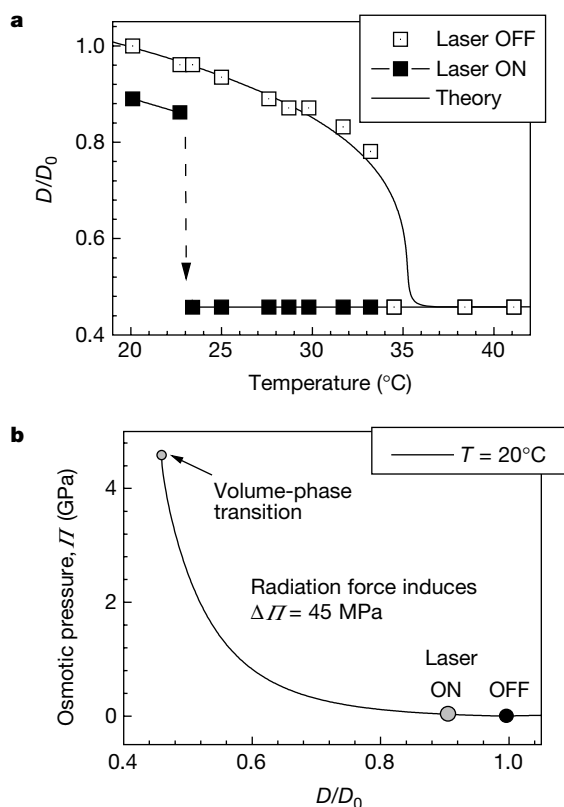


Figure 4 Volume phase transition of PNIPAM gel in D_2O . **a**, The transition induced without (open symbols) and with (filled symbols) laser illumination of 0.75 W versus temperature; the solid line shows the results of theoretical simulations. **b**, Osmotic pressure, Π , of gel at 20 °C versus diameter change, D/D_0 . The parameters used to calculate the theoretical curve were obtained by regression fitting of the experimental data in **a**. The osmotic pressure at the volume phase transition is about 4.1 GPa. Gel compaction at 20 °C induced by radiation force corresponds to $\Delta\Pi = 45$ MPa osmotic pressure (the points laser ON and OFF in **b**). The formulae used to calculate theoretical curves are given in ref. 7.

A gel collapse can be considered formally by Flory–Huggins (FH) mean-field theory, which is valid for the mixing of two phases (the swollen, and the shrunken, phases of a gel) described in terms of the change of Helmholtz free energy, ΔE , of the mixture¹⁹. FH theory describes the change of local free energy in an empirical manner by the dimensionless parameter $\chi = (\Delta H - T\Delta S)/k_B T$, which has the same sign as the free-energy change. Here ΔH and ΔS are the enthalpy and entropy changes, respectively; T is the absolute temperature and k_B stands for the Boltzmann constant. The entropy term $T\Delta S$ associated with the local motions of polymer segments can be manipulated by a radiation force, which can act on particles as small as 10 nm by hindering their local movements. The reduced mobility means a decrease in entropy, and it drives $\Delta E > 0$. Systems like PNIPAM gel, which possess a so-called upper miscibility gap¹⁹ can be described by FH theory only when the FH parameter χ is negative at low temperatures, $\Delta E < 0$, followed by an increase to positive values, $\Delta E > 0$, at higher temperatures. In this way, the volume phase transition induced by radiation force is explained as a decrease of local entropy, which induces a positive change of free energy of a mixture $\Delta E > 0$, which then becomes unstable and leads to a phase separation. Macroscopically, this is observed as a collapse of the gel due to segment–segment interaction, which causes the solvent, heavy water, to be expelled from a gel network.

When we focused laser irradiation into a diffraction-limited spot of size d ($= 1.22\lambda/\text{NA} \approx 1.5 \mu\text{m}$) with a similar axial dimension inside a gel rod of radius $a \approx 25 \mu\text{m}$ ($\text{NA} = 0.85$ is the numerical aperture of the objective lens), the illuminated gel exhibited a stimulus-sensitive response at a distance of up to 20 μm in both directions along the rod. Figure 1a shows the response on focusing the laser into the centre of the gel rod. This response is seen even though the diameter of the laser-light spot at a distance a from the focal point is less than 4 μm . This observation can be explained by considering that both bulk energy and shear energy contribute to the total energy of a gel¹⁰. The bulk energy is directly related to the volume change, which is controlled by diffusion; however, the shape deformation caused by volume changes results in the build-up of shear energy, which can be minimized by adjusting the gel's shape while keeping the volume constant²⁰. As the gels used here show a response time of about 0.1–1 s (see Supplementary Information), the scanning of a focal point at typical laser pulse rates (in the kilohertz range) might allow the recording of pre-programmed patterns and expand the range of potential gel applications to include, for example, light-sensitive actuators²⁰ and sensors. □

Received 8 May; accepted 4 September 2000.

- Schild, H. G. Poly(N-isopropylacrylamide): experiment, theory and applications. *Prog. Polym. Sci.* **17**, 163–249 (1992).
- Ilmain, F., Tanaka, T. & Kokufuta, E. Volume transition in a gel driven by hydrogen bonding. *Nature* **349**, 400–401 (1991).
- Takeoka, Y. *et al.* First order phase transition and evidence for frustrations in polyampholytic gels. *Phys. Rev. Lett.* **82**, 4863–4865 (1999).
- Annaka, M. & Tanaka, T. Multiple phases of polymer gels. *Nature* **355**, 430–432 (1992).
- Suzuki, A. & Tanaka, T. Phase transition in polymer gels induced by visible light. *Nature* **346**, 345–347 (1990).
- Hirokawa, Y. & Tanaka, T. Volume phase transition in a non-ionic gel. *J. Chem. Phys.* **81**, 6379–6380 (1984).
- Hirotsu, S., Hirokawa, Y. & Tanaka, T. Volume-phase transition of ionized N-isopropylacrylamide gels. *J. Chem. Phys.* **87**, 1392–1395 (1987).
- Smith, T. A., Hotta, J., Sasaki, K., Masuhara, H. & Itoh, Y. Photon pressure-induced association of nanometer-sized polymer chains in solutions. *J. Phys. Chem. B* **103**, 1660–1663 (1999).
- Hofkens, J., Hotta, J., Sasaki, K., Masuhara, H. & Iwai, K. Molecular assembling by the radiation pressure of a focused laser beam: poly(N-isopropylacrylamide) in aqueous solution. *Langmuir* **13**, 414–419 (1997).
- Shibayama, M. & Tanaka, T. Volume phase transition and related phenomena of polymer gels. *Adv. Polym. Sci.* **109**, 1–62 (1993).
- Svoboda, K. & Block, S. M. Biological applications of optical forces. *Annu. Rev. Biophys. Biomol. Struct.* **23**, 247–285 (1994).
- Misawa, H. & Juodkazis, S. Photophysics and photochemistry of a laser manipulated microparticle. *Prog. Polym. Sci.* **24**, 665–697 (1999).
- Juodkazis, S., Shikata, M., Takahashi, T., Matsuo, S. & Misawa, H. Fast optical switching by a laser-manipulated microdroplet of liquid crystal. *Appl. Phys. Lett.* **74**, 3627–3629 (1999).
- Ashkin, A. Acceleration and trapping of particles by radiation pressure. *Phys. Rev. Lett.* **24**, 156–159 (1970).

15. Shen, Y. R. *The Principles of Nonlinear Optics* 366–379 (Wiley & Sons, New York, 1984).
16. Ishikawa, M., Misawa, H., Kitamura, N., Fujisawa, R. & Masuhara, H. Infrared laser-induced photo-thermal phase transition of an aqueous poly(*N*-isopropylacrylamide) solution in the micrometer dimension. *Bull. Chem. Soc. Jpn* **69**, 59–66 (1996).
17. Schild, H. G. & Tirrell, D. A. Microcalorimetric detection of lower critical solution temperatures in aqueous polymer solutions. *J. Phys. Chem.* **94**, 4352–4356 (1990).
18. Fritze, M., Stern, M. B. & Wyatt, P. W. Laser-fabricated glass microlens arrays. *Opt. Lett.* **23**, 1441–143 (1998).
19. Strobl, G. R. *The Physics of Polymers* 63–128 (Springer, Berlin, 1997).
20. Hu, Z., Zang, X. & Li, Y. Synthesis and application of modulated polymer gels. *Science* **269**, 525–527 (1995).

Supplementary Information is available on Nature's World-Wide Web site (<http://www.nature.com>) or as paper copy from the London editorial office of Nature.

Acknowledgements

This present work was supported in part by a grant under Priority Area Research "Photoreaction Dynamics" from the Ministry of Education, Science, Sports and Culture of Japan, by a NEDO (New Energy Industrial Technology Development Organization) regional consortium project "Development of Ultrasensitive Micro-Integrated Analysis Systems", and by the Satellite Venture Business Laboratory of the University of Tokushima.

Correspondence and requests for materials should be addressed to H.M. (e-mail: misawa@eco.tokushima-u.ac.jp).

An enantiomerically pure hydrogen-bonded assembly

Leonard J. Prins, Feike De Jong, Peter Timmerman & David N. Reinhoudt

Laboratory of Supramolecular Chemistry and Technology, MESA⁺ Research Institute, University of Twente, PO Box 217, 7500 AE Enschede, The Netherlands

Chiral molecules have asymmetric arrangements of atoms, forming structures that are non-superposable mirror images of each other. Specific mirror images ('enantiomers') may be obtained either from enantiomerically pure precursor compounds, through enantioselective synthesis, or by resolution of so-called racemic mixtures of opposite enantiomers, provided that racemization (the spontaneous interconversion of enantiomers) is sufficiently slow. Non-covalent assemblies can similarly adopt chiral supramolecular structures^{1,2}, and if they are held together by relatively strong interactions, such as metal coordination³, methods analogous to those used to obtain chiral molecules yield enantiomerically pure non-covalent products. But the resolution of assemblies formed through weak interactions, such as hydrogen-bonding, remains challenging, reflecting their lower stability and significantly higher susceptibility to racemization. Here we report the design of supramolecular structures from achiral calix[4]arene dimelamines and cyanurates, which form multiple cooperative hydrogen bonds that together provide sufficient stability to allow the isolation of enantiomerically pure assemblies. Our design strategy is based on a non-covalent 'chiral memory' concept^{4,5}, whereby we first use chiral barbiturates to induce the supramolecular chirality in a hydrogen-bonded assembly⁶, and then substitute them by achiral cyanurates. The stability of the resultant chiral assemblies in benzene, a non-polar solvent not competing for hydrogen bonds, is manifested by a half-life to racemization of more than four days at room temperature.

There is considerable interest in hydrogen-bonded assemblies that display chirality at the supramolecular level^{7–10}. Most studies have focused on the induction of supramolecular chirality by means of the introduction of chiral centres (*R* or *S*) in the components: the two chiral forms (for instance *M* and *P*) of the assembly are then no longer mirror images, but instead have a diastereomeric relationship. Both diastereomers are usually present in unequal amounts,

because they have different free energies. The diastereomeric excess, d.e. (d.e. = $100(M - P)/(M + P)$) is directly related to the difference in free energy at thermodynamic equilibrium.

Hydrogen-bonded assemblies such as $I_3 \cdot (CYA)_6$ (Fig. 1) exist exclusively as a single diastereomer (d.e. > 98%) when they are composed of chiral dimelamines or cyanurates^{6,11} (Fig. 2a). With achiral components, the assemblies are formed as an equal mixture of the (*M*)- and (*P*)-enantiomers^{12,13} (Fig. 2b). The energy barrier for racemization is presumably quite high, because the lowest-energy pathway for racemization involves the breakage of at least six hydrogen bonds. Consequently, the single enantiomers should be kinetically stable enough at room temperature to be isolated in optically pure form. However, enantiomers by definition have identical free energies, which precludes their selective synthesis without an external chiral bias. Therefore, their synthesis is conducted via previous induction of the chirality using chiral barbiturates, followed by substitution of the chiral components for non-chiral cyanurates.

The assembly of calix[4]arene dimelamine **1** and chiral barbiturate (*R*)-BAR (see Fig. 1) in benzene occurs with complete diastereoselectivity, and gives exclusively the (*M*)-diastereomer of assembly $I_3 \cdot ((R)\text{-BAR})_6$ (Fig. 3a). ¹H NMR spectroscopy confirms the complete induction of chirality (d.e. > 98%), as only two signals at 12.69 and 14.46 p.p.m. (for the two magnetically inequivalent N-H protons of (*R*)-BAR) are present (Fig. 3c, trace I). Signals for the corresponding (*P*)-diastereomer of assembly $I_3 \cdot ((R)\text{-BAR})_6$ are not observed. In the same way, assembly of dimelamine **1** with (*S*)-BAR gives exclusively the (*P*)-diastereomer of $I_3 \cdot ((S)\text{-BAR})_6$. The assemblies (*M*)- $I_3 \cdot ((R)\text{-BAR})_6$ and (*P*)- $I_3 \cdot ((S)\text{-BAR})_6$ are enantiomers, and their circular dichroism (CD) spectra—which are highly characteristic for calix[4]arene double rosette assemblies—reflect their opposite handedness (Fig. 3b). Based on previously studied assemblies, we attribute a positive CD curve to (*M*)-chirality and a negative curve to (*P*)-chirality⁶.

The chiral barbiturates within (*M*)- $I_3 \cdot ((R)\text{-BAR})_6$ can then be substituted by achiral cyanurates. Addition of achiral BuCYA to (*M*)- $I_3 \cdot ((R)\text{-BAR})_6$ in benzene-*d*₆ (as studied by ¹H NMR spectroscopy) reveals that exchange of (*R*)-BAR for BuCYA occurs instantaneously (Fig. 3c). At a 1:1.2 ratio of (*R*)-BAR and BuCYA, the exchange is quantitative, as proven by the new set of signals at 14.42 and 14.89 p.p.m. for assembly $I_3 \cdot (\text{BuCYA})_6$ and the complete disappearance of the signals at 12.69 and 14.46 p.p.m. for assembly (*M*)- $I_3 \cdot ((R)\text{-BAR})_6$ (Fig. 3c trace V). Additional evidence for the quantitative exchange of barbiturates for cyanurates was obtained from MALDI-TOF (matrix assisted laser desorption/ionization-time of flight) mass spectrometry using Ag⁺-labelling¹⁴ (data not shown). The mass spectrum after addition of 2.3 equivalents of BuCYA to Ag⁺-labelled assembly $3_3 \cdot (\text{DEB})_6$ showed only a signal at a mass-to-charge ratio of 4,487.3, corresponding to assembly $3_3 \cdot (\text{BuCYA})_6$. Signals for any of the other assemblies $3_3 \cdot (\text{DEB})_n(\text{BuCYA})_{6-n}$ (*n* = 1–6) were not observed. The barbiturate–cyanurate exchange process occurs because cyanurates form much stronger hydrogen bonds with melamines than barbiturates, as reflected by the association constants of $(1.0 \pm 0.1) \times 10^2$ and $(2.2 \pm 0.8) \times 10^3 \text{ M}^{-1}$ in CDCl₃ for model complexes **4-5** and **4-6**, respectively. The increased affinity of melamines for cyanurates is probably the result of the increased acidity of the cyanurate N-H protons^{15,16}.

The newly formed assembly $I_3 \cdot (\text{BuCYA})_6$ still has a very intense CD spectrum despite the fact that it no longer contains any chiral components (Fig. 3b). This shows that the assembly has preserved the (*M*)-chirality of assembly $I_3 \cdot ((R)\text{-BAR})_6$ as induced by (*R*)-BAR. Independent ¹H NMR exchange experiments with assemblies $I_3 \cdot (\text{DEB})_6$ and $2_3 \cdot (\text{DEB})_6$ proved that no exchange of the dimelamine components occurs under the conditions for barbiturate/cyanurate exchange, and thus no racemization of the assemblies occurs. Moreover, the fact that the observed CD intensity for

(*M*)-**1**₃·(BuCYA)₆ ($\Delta\epsilon_{\text{max}} \approx 90 \text{ cm}^2 \text{ mmol}^{-1}$) is similar to that of the pure diastereomers of similar assemblies ($\Delta\epsilon_{\text{max}} \approx 80\text{--}100 \text{ cm}^2 \text{ mmol}^{-1}$)⁶ provides strong evidence for the fact that assembly **1**₃·(BuCYA)₆ is exclusively present as the *M*-enantiomer, with an enantiomeric excess (e.e.) that is comparable to the d.e. of assembly (*M*)-**1**₃·((*R*)-BAR)₆—that is, > 98%. Similarly, the (*P*)-enantiomer of assembly **1**₃·(BuCYA)₆ could be obtained by addition of BuCYA to diastereomer (*P*)-**1**₃·((*S*)-BAR)₆, as reflected by the opposite sign of the CD curve compared to assembly (*M*)-**1**₃·(BuCYA)₆ (Fig. 3b).

Several control experiments were performed to exclude the possibility that the memory of chirality observed in **1**₃·(BuCYA)₆ is due to the presence of free (*R*)-BAR (refs 17–19). We found that addition of free (*R*)-BAR to racemic assembly (*M/P*)-**1**₃·(BuCYA)₆ does not lead to a CD signal of significant intensity ($\Delta\epsilon_{\text{max}} < 5 \text{ cm}^2 \text{ mmol}^{-1}$), even after standing for one hour

at room temperature. Moreover, the CD intensity of assembly (*P*)-**1**₃·(BuCYA)₆ in the presence of the free (*S*)-BAR is reduced to zero after heating a benzene solution at 70 °C for 45 min, and is not restored after cooling the solution to room temperature and standing for an additional hour.

The racemization rate was measured by monitoring the decrease with time of the CD intensity at different temperatures (Fig. 4a). Analysis of these data using Arrhenius' equation gave an activation energy of 105.9 kJ mol⁻¹, with a corresponding half-life to racemization (defined as the time required for a 50% decrease in e.e.) of approximately 4.5 d at room temperature (Fig. 4b). The large half-life time reflects the high kinetic stability of the enantiomers of this hydrogen-bonded assembly.

The enantioselective formation of the (*P*)- and (*M*)-enantiomers of assembly **1**₃·(BuCYA)₆ provides an opportunity to study the mechanism of the racemization process (*P*)-**1**₃·(BuCYA)₆ ⇌ (*M*)-**1**₃·(BuCYA)₆. Detailed kinetic studies revealed that racemization of assembly (*P*)-**1**₃·(BuCYA)₆ can proceed via both an uncatalysed and a catalysed pathway, with the liberated (*S*)-BAR as the catalyst (see Supplementary Information). Both processes are first order in assembly (*P*)-**1**₃·(BuCYA)₆. The initial rate of racemization can be expressed using the following equation:

$$R_{t=0} = \frac{d[(P)\text{-}\mathbf{1}_3 \cdot (\text{BuCYA})_6]}{dt} = (k_{\text{cat}}[(S)\text{-BAR}] + k_{\text{uncat}})[(P)\text{-}\mathbf{1}_3 \cdot (\text{BuCYA})_6]_0$$

with $k_{\text{cat}} = 7.4 \times 10^{-3} \text{ l mol}^{-1} \text{ s}^{-1}$ and $k_{\text{uncat}} = 1.1 \times 10^{-5} \text{ s}^{-1}$ (50 °C). (*S*)-BAR probably acts as a catalyst via the formation of an activated complex with (*P*)-**1**₃·(BuCYA)₆, the racemization of which is much faster. The very low solubility of free BuCYA in benzene (< 10⁻⁴ mM) precludes a similar catalytic role for this compound.

The kinetic data do not indicate whether dissociation of the assembly is a prerequisite for racemization to occur, as an

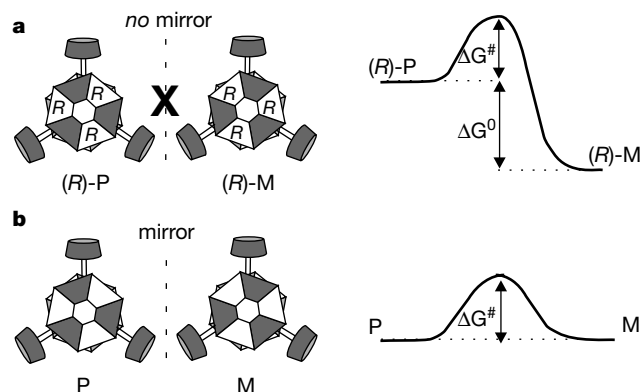


Figure 2 Schematic representations of diastereomeric (a) and enantiomeric (b) assemblies and their difference in free energy.

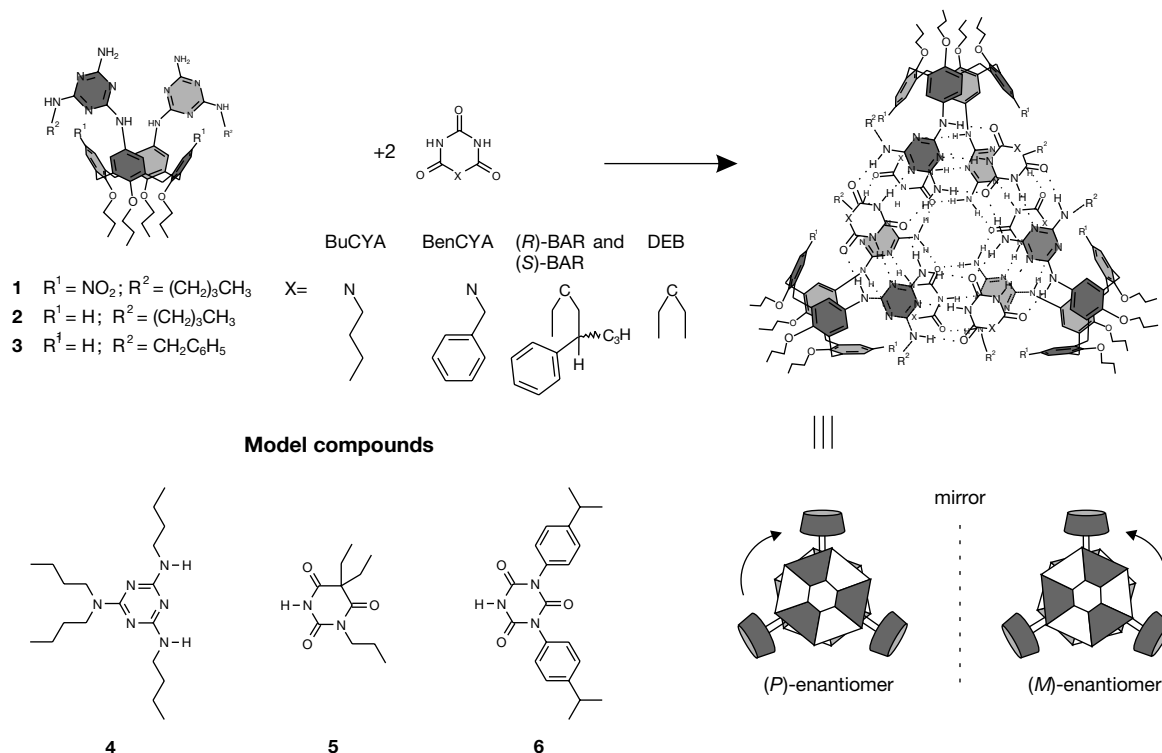


Figure 1 Formation of non-covalent chiral assemblies with typical composition **1**₃·(BAR)₆ and **1**₃·(CYA)₆. Only the assembly with D₃-symmetry is shown (top right), with schematic representations for both the (*M*)- and (*P*)-enantiomers (bottom right). The assignment of

the *M/P*-configuration is based on a clockwise (*P*) or anticlockwise (*M*) orientation of the two melamine units of **1** in assembly **1**₃·(BAR)₆. Model compounds **4–6** were used to study the difference in association constants between barbiturates and cyanurates.

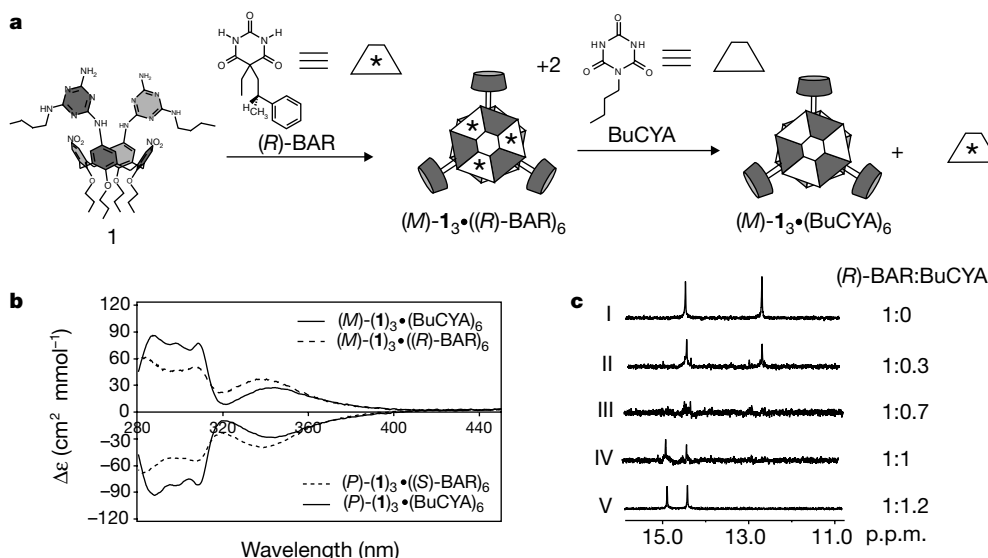


Figure 3 Non-covalent synthesis of an enantiomerically pure hydrogen-bonded assembly. **a**, Enantioselective synthesis of $(M)\text{-}1_3\text{•}(\text{BuCYA})_6$; **b**, the CD spectra of assemblies $(M)\text{-}1_3\text{•}((R)\text{-BAR})_6$ (bold dashed line), $(M)\text{-}1_3\text{•}(\text{BuCYA})_6$ (bold solid line), $(P)\text{-}1_3\text{•}((S)\text{-BAR})_6$ (dashed line), and $(P)\text{-}1_3\text{•}(\text{BuCYA})_6$ (solid line); **c**, a ^1H NMR (300 MHz)

titration study of $(M)\text{-}1_3\text{•}((R)\text{-BAR})_6$ (trace I) with BuCYA in benzene- d_6 (traces II–IV), resulting in the quantitative formation of assembly $(M)\text{-}1_3\text{•}(\text{BuCYA})_6$ (trace V) at a 1:1.2 ratio of $(R)\text{-BAR}$ to BuCYA. All spectra were recorded in benzene at 1 mM assembly concentrations.

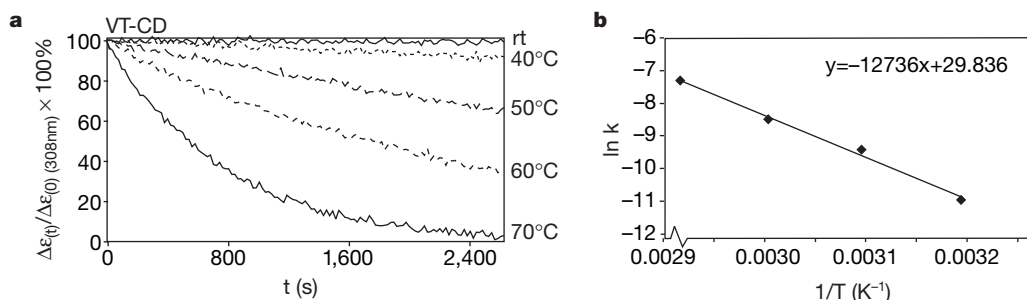


Figure 4 Study of the racemization process by circular dichroism. **a**, Temperature-dependent racemization of assembly $(P)\text{-}1_3\text{•}(\text{BuCYA})_6$ followed by the decrease of the CD intensity at 308 nm, and **b**, analysis of the kinetic data using an Arrhenius plot.

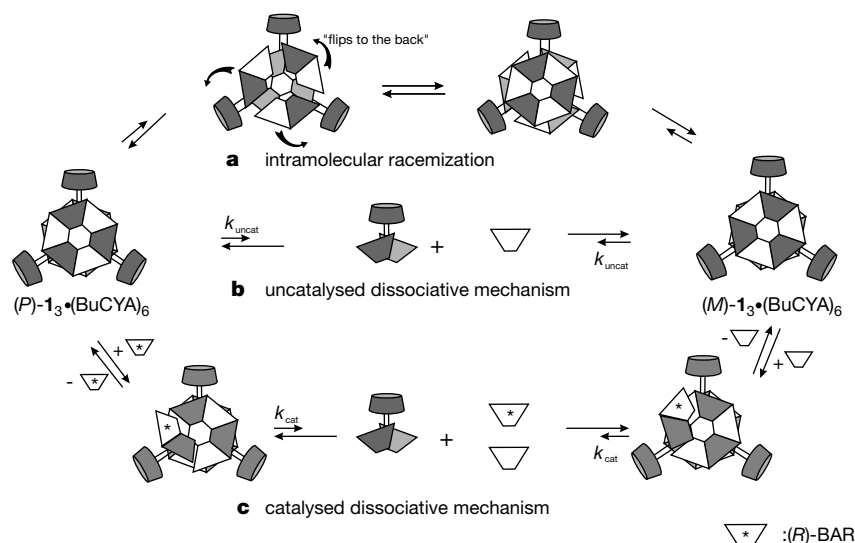


Figure 5 Schematic representations of the different racemization pathways. The figure shows the racemization of assembly $(M)\text{-}1_3\text{•}(\text{BuCYA})_6$ via: **a**, an intramolecular mechanism; and **b**, an uncatalysed or **c**, a catalysed, dissociative mechanism.

intramolecular mechanism is also feasible (Fig. 5a). We obtained strong evidence for a dissociative mechanism by measuring the rate of dimelamine exchange between assemblies $1_3 \cdot (\text{BuCYA})_6$ and $2_3 \cdot (\text{BuCYA})_6$ by ^1H NMR spectroscopy. Upon mixing these assemblies in a 1:1 ratio in benzene- d_6 at 70°C , well-separated signals for the new heteromeric assemblies $1_2 2_1 \cdot (\text{BuCYA})_6$ and $1_1 2_2 \cdot (\text{BuCYA})_6$ appear over time as a result of the exchange of calix[4]arene dimelamines 1 and 2. The rate-determining step in the exchange process involves the dissociation of calix[4]arene dimelamines from an intact assembly, a process that requires the disruption of a total of 12 hydrogen bonds. If racemization occurred via a dissociative mechanism, the rates of dimelamine exchange and racemization should be of similar magnitude. We studied the concentration dependence of the rate of dimelamine exchange, both with and without (S)-BAR present, and found that this rate can be expressed by the same equation as the rate of racemization, with slightly different values of k_{cat} and k_{uncat} ($14.7 \times 10^{-3} \text{ l mol}^{-1} \text{ s}^{-1}$ and $7.0 \times 10^{-5} \text{ s}^{-1}$, respectively, at 70°C). These values relate well to those obtained from the racemization studies. (The slightly higher values for dimelamine exchange are due to the slightly higher temperature at which these ^1H NMR experiments were conducted, 70 versus 50°C). From these results we conclude that racemization occurs predominantly via a dissociative mechanism, both uncatalysed and catalysed (Fig. 5b and c). The rate-limiting step involves the dissociation of calix[4]arene dimelamines from the intact assembly, followed by rapid racemization and formation of the opposite enantiomer. □

Received 17 March; accepted 4 September 2000.

- Lehn, J.-M. *Supramolecular Chemistry, Concepts and Perspectives* (VCH, Weinheim, 1995).
- Whitesides, G. M. *et al.* Noncovalent synthesis: using physical-organic chemistry to make aggregates. *Acc. Chem. Res.* **28**, 37–44 (1995).
- Piguet, C., Bernardinelli, G. & Hopfgartner, G. Helicates as versatile supramolecular complexes. *Chem. Rev.* **97**, 2005–2062 (1997).
- Yashima, E., Maeda, K. & Okamoto, Y. Memory of macromolecular helicity assisted by interaction with achiral small molecules. *Nature* **399**, 449–451 (1999).
- Furusho, Y., Kimura, T., Mizuno, Y. & Aida, T. Chirality-memory molecule: a D_2 -symmetric fully substituted porphyrin as a conceptually new chirality sensor. *J. Am. Chem. Soc.* **119**, 5267–5268 (1997).
- Prins, L. J., Huskens, J., De Jong, F., Timmerman, P. & Reinhoudt, D. N. Complete asymmetric induction of supramolecular chirality in a hydrogen-bonded assembly. *Nature* **398**, 498–502 (1999).
- Suárez, M., Branda, N., Lehn, J.-M., Decian, A. & Fischer, J. Supramolecular chirality: chiral hydrogen-bonded supermolecules from achiral molecular components. *Helv. Chim. Acta* **81**, 1–13 (1998).
- Rivera, J. M., Martin, T. & Rebek, J. Jr Chiral spaces: dissymmetric capsules through self-assembly. *Science* **279**, 1021–1023 (1998).
- Castellano, R. K., Nuckolls, C. & Rebek, J. Jr Transfer of chiral information through molecular assembly. *J. Am. Chem. Soc.* **121**, 11156–11163 (1999).
- Qiao, S., Choi, I. S. & Whitesides, G. M. Observation of diastereomers of the hydrogen-bonded aggregate $\text{Hub}(\text{M})_3 \cdot 3\text{CA}$ using ^1H NMR spectroscopy when CA is an optically-active isocyanuric acid. *J. Org. Chem.* **62**, 2619–2621 (1997).
- Prins, L. J., Jolliffe, K. A., Hulst, R., Timmerman, P. & Reinhoudt, D. N. Control of structural isomerism in noncovalent hydrogen bonded assemblies using peripheral chiral information. *J. Am. Chem. Soc.* **122**, 3617–3627 (2000).
- Vreekamp, R. H., Van Duynhoven, J. P. M., Hubert, M., Verboom, W. & Reinhoudt, D. N. Molecular boxes based on calix[4]arene double rosettes. *Angew. Chem. Int. Edn Engl.* **35**, 1215–1218 (1996).
- Timmerman, P. *et al.* Noncovalent assembly of functional groups on calix[4]arene molecular boxes. *Chem. Eur. J.* **3**, 1823–1832 (1997).
- Jolliffe, K. A. *et al.* Characterization of supramolecular hydrogen-bonded assemblies by MALDI-TOF mass spectrometry after Ag^+ -labelling. *Angew. Chem. Int. Edn Engl.* **37**, 1294–1297 (1998).
- Mascal, M. *et al.* The ion-pair reinforced, hydrogen-bonding molecular ribbon. *J. Chem. Soc. Chem. Commun.* 805–806 (1995).
- Shieh, H. S. & Voet, D. The x-ray structure of the molecular complex 8-bromo-9-ethyladenine-cyanuric acid monohydrate. *Acta Crystallogr. B* **32**, 2354–2360 (1976).
- Yashima, E., Matsumura, T. & Okamoto, Y. Chirality assignment of amines and amino alcohols based on circular dichroism induced by helix formation of a stereoregular poly((4-carboxyphenyl)acetylene) through acid-base complexation. *J. Am. Chem. Soc.* **119**, 6345–6359 (1997).
- Palmans, A. R. A., Vekemans, J. A. J. M., Havinga, E. E. & Meijer, E. W. Sergeants-and-soldiers principle in chiral columnar stacks of disc-shaped molecules with C_3 symmetry. *Angew. Chem. Int. Edn Engl.* **36**, 2648–2651 (1997).
- Green, M. M. *et al.* A helical polymer with a cooperative response to chiral information. *Science* **268**, 1860–1866 (1995).

Supplementary Information is available on Nature's World-Wide Web site (<http://www.nature.com>) or as paper copy from the London editorial office of Nature.

Acknowledgements

We thank C. Marjo and A. Bielejewska for their work on the model complexes 4-5 and 4-6, and R. Fokkens and N.M.M. Nibbering for the MALDI-TOF MS measurements. This work was supported by the Council for Chemical Sciences of the Netherlands Organization for Scientific Research (CW-NWO).

Correspondence and requests for materials should be addressed to D.N.R. (e-mail: smct@ct.utwente.nl).

Acceleration of global warming due to carbon-cycle feedbacks in a coupled climate model

Peter M. Cox*, Richard A. Betts*, Chris D. Jones*, Steven A. Spall* & Ian J. Totterdell†

* Hadley Centre, The Met Office, Bracknell, Berkshire RG12 2SY, UK

† Southampton Oceanography Centre, European Way, Southampton SO14 3ZH, UK

The continued increase in the atmospheric concentration of carbon dioxide due to anthropogenic emissions is predicted to lead to significant changes in climate¹. About half of the current emissions are being absorbed by the ocean and by land ecosystems², but this absorption is sensitive to climate^{3,4} as well as to atmospheric carbon dioxide concentrations⁵, creating a feedback loop. General circulation models have generally excluded the feedback between climate and the biosphere, using static vegetation distributions and CO_2 concentrations from simple carbon-cycle models that do not include climate change⁶. Here we present results from a fully coupled, three-dimensional carbon-climate model, indicating that carbon-cycle feedbacks could significantly accelerate climate change over the twenty-first century. We find that under a 'business as usual' scenario, the terrestrial biosphere acts as an overall carbon sink until about 2050, but turns into a source thereafter. By 2100, the ocean uptake rate of 5 Gt C yr^{-1} is balanced by the terrestrial carbon source, and atmospheric CO_2 concentrations are 250 p.p.m.v. higher in our fully coupled simulation than in uncoupled carbon models², resulting in a global-mean warming of 5.5 K, as compared to 4 K without the carbon-cycle feedback.

The general circulation model (GCM) that we used is based on the third Hadley Centre coupled ocean-atmosphere model, HadCM3⁷, which we have coupled to an ocean carbon-cycle model (HadOCC) and a dynamic global vegetation model (TRIFFID). The atmospheric physics and dynamics of our GCM are identical to those used in HadCM3, but the additional computational expense of including an interactive carbon cycle made it necessary to reduce the ocean resolution to $2.5^\circ \times 3.75^\circ$, necessitating the use of flux adjustments in the ocean component to counteract climate drift. HadOCC accounts for the atmosphere-ocean exchange of CO_2 , and the transfer of CO_2 to depth through both the solubility pump and the biological pump⁸. TRIFFID models the state of the biosphere in terms of the soil carbon, and the structure and coverage of five functional types of plant within each model gridbox (broadleaf tree, needleleaf tree, C_3 grass, C_4 grass and shrub). Further details on HadOCC and TRIFFID are given in Methods.

The coupled climate/carbon-cycle model was brought to equilibrium with a 'pre-industrial' atmospheric CO_2 concentration of 290 p.p.m.v., starting from an observed landcover data set⁹. The resulting state was stable, with negligible net land-atmosphere and

ocean–atmosphere carbon fluxes in the long-term mean, and no discernible drift in atmospheric CO₂ concentration. This simulation produces the locations of the main land biomes, and estimates of ocean carbon (38,100 Gt C), vegetation carbon (493 Gt C), soil carbon (1,180 Gt C) and terrestrial net primary productivity (60 Gt C yr⁻¹) that are within the range of other estimates^{2,10–12}. Ocean primary productivity is also compatible with results derived from remote sensing^{13,14}, producing a global-mean total of 53 Gt C yr⁻¹, and realistic seasonal and latitudinal variations¹⁵.

The simulated carbon cycle displays significant interannual variability, which is driven by the model-generated El Niño/Southern Oscillation (ENSO). A realistic response to internal climate variability is an important prerequisite for any carbon-cycle model to be used in climate change predictions. Fluctuations in annual-mean atmospheric CO₂ are correlated with the phase of ENSO, as indicated by the Nino3 index (Fig. 1). During El Niño conditions (positive Nino3), the model simulates an increase in atmospheric CO₂; this increase results from the terrestrial biosphere acting as a large source (especially in Amazonia¹⁶), which is only partially offset by a reduced outgassing from the tropical Pacific Ocean. The opposite is true during the La Niña phase. The overall sensitivity of the modelled carbon cycle to ENSO variability is consistent with the observational record¹⁷, demonstrating that the coupled system responds realistically to climate anomalies.

Transient simulations were carried out for 1860–2100, using CO₂ emissions as given by the IS92a scenario¹⁸. Other greenhouse gases were also prescribed from IS92a, but the radiative effects of sulphate aerosols were omitted. Three separate runs were completed to isolate the effect of climate/carbon-cycle feedbacks; an experiment with prescribed IS92a CO₂ and fixed vegetation (that is, a ‘standard’ GCM climate change simulation), an experiment with interactive CO₂ and dynamic vegetation but no direct effects of CO₂ on climate (akin to ‘offline’ carbon-cycle projections that neglect climate change⁶), and a fully coupled climate/carbon-cycle simulation.

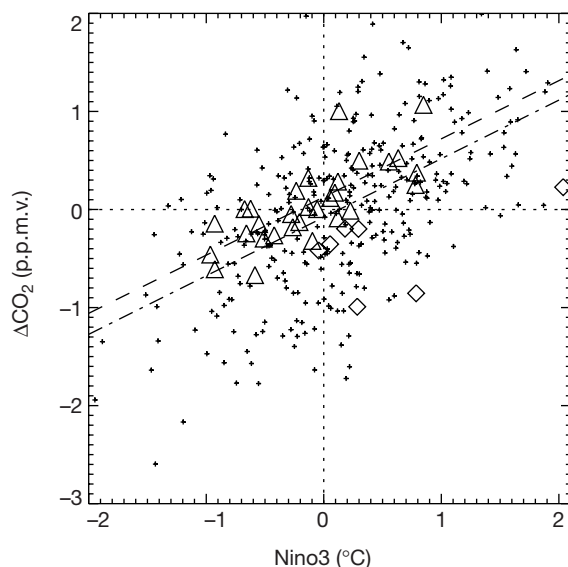


Figure 1 Modelled and observed interannual variability in the atmospheric CO₂ concentration. The figure shows the anomaly in the growth rate of atmospheric CO₂ versus the Nino3 index, taken from our pre-industrial control simulation (crosses) and the Mauna Loa observations (triangles). (The Nino3 index is the annual mean sea surface temperature anomaly in the tropical Pacific, 150° W–90° W, 5° S–5° N.) The gradients of the dashed and dot-dashed lines represent the sensitivity of the carbon cycle to ENSO, as given by the observations and the model, respectively. We have excluded observations that immediately follow major volcanic events (data points shown by diamonds), since during these years the carbon cycle may have been significantly perturbed by the induced tropospheric cooling.

Figure 2 shows results from the fully coupled run. From 1860 to 2000, the simulated stores of carbon in the ocean and on land increase by about 100 Gt C and 75 Gt C, respectively. However, the atmospheric CO₂ is 15–20 p.p.m.v. too high by the present day (corresponding to a timing error of about 10 years). Possible reasons for this include an overestimate of the prescribed net land-use emissions and the absence of other important climate forcing factors. The modelled global mean temperature increase from 1860 to 2000 is about 1.4 K (Fig. 3b), which is higher than observed¹⁹, probably due to the absence of cooling from anthropogenic aerosols²⁰. Offline tests suggest that such a relative warming can suppress the terrestrial carbon sink by enhancing soil and plant respiration¹¹. Nevertheless, the rate of increase of CO₂ from 1950 to 2000 closely follows the recent observational record, which implies that the airborne fraction is being well simulated over this period. For the 20 years centred on 1985, the mean land and ocean uptake of carbon are 1.5 and 1.6 Gt yr⁻¹, respectively (compare best estimates for the 1980s of 1.8 ± 1.8 and 2.0 ± 0.8 Gt yr⁻¹)². The model therefore captures the most important characteristics of the present-day carbon cycle.

The simulated atmospheric CO₂ diverges much more rapidly from the standard IS92a concentration scenario in the future. First, vegetation carbon in South America begins to decline, as a drying and warming of Amazonia initiates loss of forest (Fig. 4a). This is driven purely by climate change, as can be seen by comparing the fully coupled run (red lines) to the run without global warming (blue lines). The effects of anthropogenic deforestation on land-cover are neglected in both cases. A second critical point is reached at about 2050, when the land biosphere as a whole switches from being a weak sink for CO₂ to being a strong source (Fig. 2). The reduction in terrestrial carbon from around 2050 onward is associated with a widespread climate-driven loss of soil carbon (Fig. 4b). An increase in the concentration of atmospheric CO₂ alone tends to increase the rate of photosynthesis and thus terrestrial carbon storage, provided that other resources are not limiting⁴. However, plant maintenance and soil respiration rates both increase with temperature. As a consequence, climate warming (the indirect effect of a CO₂ increase) tends to reduce terrestrial carbon storage¹¹, especially in the warmer regions where an increase in temperature is

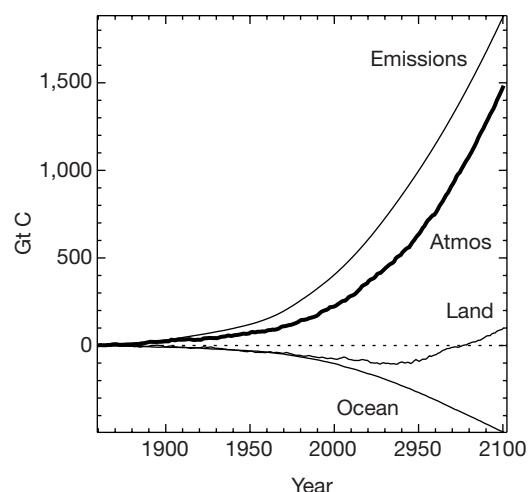


Figure 2 Budgets of carbon during the fully coupled simulation. The thick line shows the simulated change in atmospheric CO₂. The thinner lines show the integrated impact of the emissions, and of land and ocean fluxes, on the atmospheric CO₂ increase, with negative values implying net uptake of CO₂. We note that the terrestrial biosphere takes up CO₂ at a decreasing rate from about 2010 onwards, becoming a net source at around 2050. By 2100 this source from the land almost balances the oceanic sink, so that atmospheric carbon content is increasing at about the same rate as the integrated emissions (that is, the airborne fraction is ~1).

not beneficial for photosynthesis. At low CO_2 concentrations the direct effect of CO_2 dominates, and both vegetation and soil carbon increase with atmospheric CO_2 . But as CO_2 increases further, terrestrial carbon begins to decrease, because the direct effect of CO_2 on photosynthesis saturates but the specific soil respiration rate continues to increase with temperature. The transition between these two regimes occurs abruptly at around 2050 in this experiment (Fig. 4b). The carbon stored on land decreases by about 170 Gt C from 2000 to 2100, accelerating the rate of atmospheric CO_2 increase over this period.

The ocean takes up about 400 Gt C over the same period, but at a rate which is asymptotically approaching 5 Gt C yr^{-1} by 2100. This reduced efficiency of oceanic uptake is partly a consequence of the nonlinear dependence of the partial pressure of dissolved CO_2 on the total ocean carbon concentration, but may also have contributions from climate change³. Although the thermohaline circulation of the ocean weakens²¹ by about 25% from 2000 to 2100, this is much less of a reduction than seen in some previous simulations²², and the corresponding effect on ocean carbon uptake is less significant. In our experiment, increased thermal stratification due to warming of the sea surface suppresses upwelling, which reduces nutrient availability and lowers primary production by about 5%. However, ocean-only tests suggest a small effect of climate change on oceanic carbon uptake, as this reduction in the

biological pump is compensated by reduced upwelling of deep waters which have high concentrations of total carbon.

By 2100 the modelled CO_2 concentration is about 980 p.p.m.v. in the coupled experiment, which is more than 250 p.p.m.v. higher than the standard IS92a scenario or that simulated in the 'offline' experiment (Fig. 3a). As a result, the global-mean land temperatures increase from 1860 to 2100 by about 8 K, rather than the 5.5 K of the standard scenario (Fig. 3b).

These numerical experiments demonstrate the potential importance of climate/carbon-cycle feedbacks, but the magnitude of these in the real Earth system is still highly uncertain. Terrestrial carbon models differ in their responses to climate change^{11,12}, owing to gaps in basic understanding of processes. For example, the potential conversion of the global terrestrial carbon sink to a source is critically dependent upon the long-term sensitivity of soil respiration to global warming, which is still a subject of debate²³. The experiments presented here exclude the potentially large direct human influences on terrestrial carbon uptake through changes in landcover and land management. Local effects, such as the possible climate-driven loss of the Amazon rainforest, rest upon uncertain aspects of regional climate change, and may be 'short-circuited' by direct human deforestation. A full assessment of the uncertainties must await further coupled experiments utilizing alternative representations of processes and including a more complete set of natural

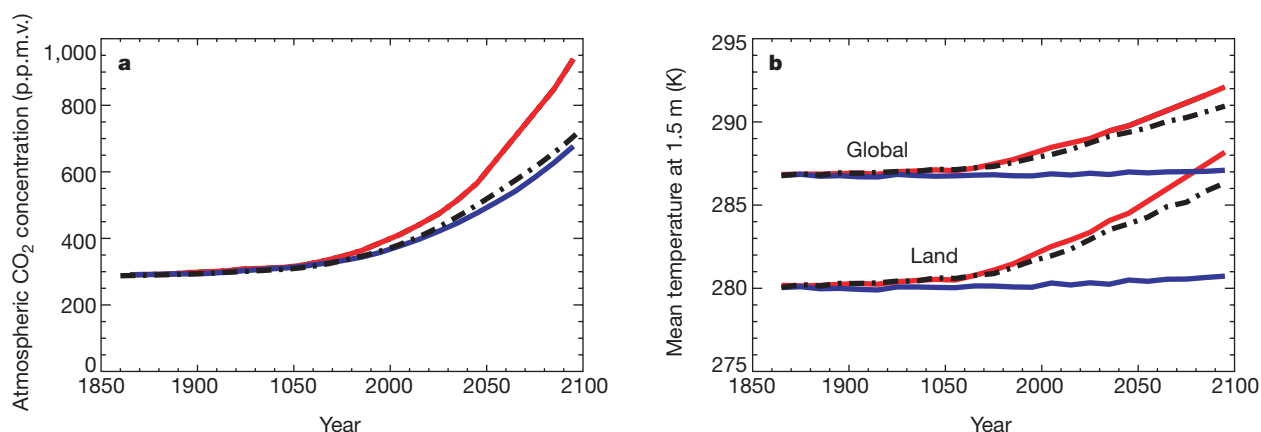


Figure 3 Effect of climate/carbon-cycle feedbacks on CO_2 increase and global warming. **a**, Global-mean CO_2 concentration, and **b**, global-mean and land-mean temperature, versus year. Three simulations are shown; the fully coupled simulation with interactive CO_2 and dynamic vegetation (red lines), a standard GCM climate change simulation with

prescribed (IS92a) CO_2 concentration and fixed vegetation (dot-dashed lines), and the simulation which neglects direct CO_2 -induced climate change (blue lines). The slight warming in the latter is due to CO_2 -induced changes in stomatal conductance and vegetation distribution.

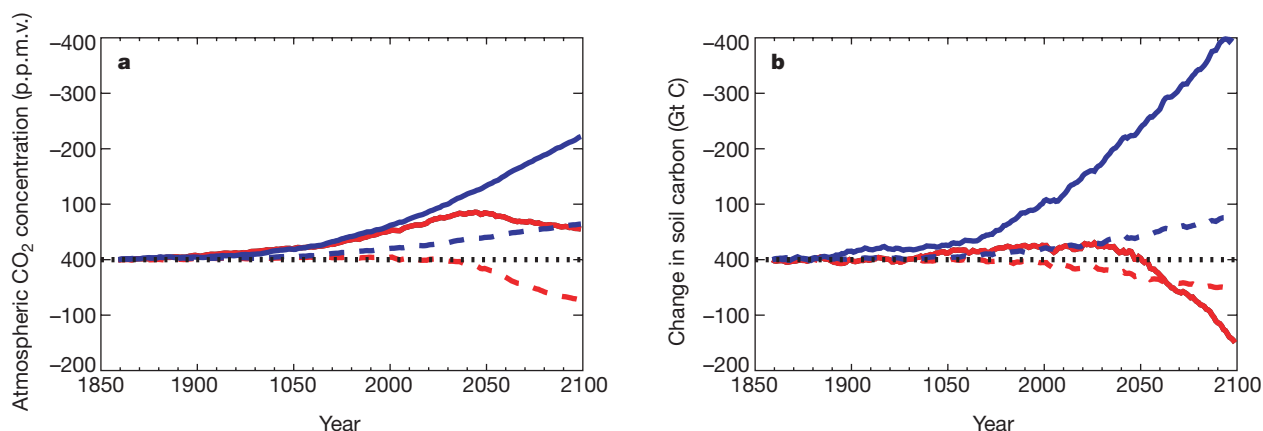


Figure 4 Effect of global warming on changes in land carbon storage. The red lines represent the fully coupled climate/carbon-cycle simulation, and the blue lines are from the 'offline' simulation which neglects direct CO_2 -induced climate change. The figure

shows simulated changes in vegetation carbon (**a**) and soil carbon (**b**) for the global land area (continuous lines) and South America alone (dashed lines).

and anthropogenic forcing factors (for example, land-use change, forest fires, sulphate aerosol concentrations and nitrogen deposition). However, our results indicate that it will be essential to accurately represent previously neglected carbon-cycle feedbacks if we are to successfully predict climate change over the next 100 years. □

Methods

Ocean carbon-cycle model

The inorganic component of HadOCC has been extensively tested as part of the Ocean Carbon Cycle Intercomparison Project; it was found to reproduce tracer distributions to an accuracy consistent with other ocean GCMs²⁴. The biological component treats four additional ocean fields: nutrient, phytoplankton, zooplankton and detritus⁸. The phytoplankton population changes as a result of the balance between growth, which is controlled by light level and the local concentration of nutrient, and mortality, which is mostly as a result of grazing by zooplankton. Detritus, which is formed by zooplankton excretion and by phyto- and zooplankton mortality, sinks at a fixed rate and slowly remineralizes to reform nutrient and dissolved inorganic carbon. Thus both nutrient and carbon are absorbed by phytoplankton near the ocean surface, pass up the food chain to zooplankton, and are eventually remineralized from detritus in the deeper ocean. The model also includes the formation of calcium carbonate and its dissolution at depth (below the lysocline).

Terrestrial carbon-cycle model

TRIFFID (top-down representation of interactive foliage and flora including dynamics) has been used offline in a comparison of dynamic global vegetation models¹¹. Carbon fluxes for each vegetation type are calculated every 30 minutes as a function of climate and atmospheric CO₂ concentration, from a coupled photosynthesis/stomatal-conductance scheme^{25,26}, which utilizes existing models of leaf-level photosynthesis in C₃ and C₄ plants^{27,28}. The accumulated fluxes are used to update the vegetation and soil carbon every 10 days. The natural landcover evolves dynamically based on competition between the vegetation types, which is modelled using a Lotka–Volterra approach and a tree–shrub–grass dominance hierarchy. We also prescribe some agricultural regions, in which grasslands are assumed to be dominant. Carbon lost from the vegetation as a result of local litterfall or large-scale disturbance is transferred into a soil carbon pool, where it is broken down by microorganisms that return CO₂ to the atmosphere. The soil respiration rate is assumed to double for every 10 K of warming²⁹, and is also dependent on the soil moisture content³⁰. Changes in the biophysical properties of the land surface⁵, as well as changes in terrestrial carbon, feed back onto the atmosphere.

Received 6 January; accepted 26 September 2000.

- Houghton, J. T. *et al.* (eds) *Climate Change 1995: The Science of Climate Change* (Cambridge Univ. Press, Cambridge, 1996).
- Schimel, D. *et al.* in *Climate Change 1995: The Science of Climate Change* Ch. 2 (eds Houghton, J. T. *et al.*) 65–131 (Cambridge Univ. Press, Cambridge, 1995).
- Sarmiento, J., Hughes, T., Stouffer, R. & Manabe, S. Simulated response of the ocean carbon cycle to anthropogenic climate warming. *Nature* **393**, 245–249 (1998).
- Cao, M. & Woodward, F. I. Dynamic responses of terrestrial ecosystem carbon cycling to global climate change. *Nature* **393**, 249–252 (1998).
- Betts, R. A., Cox, P. M., Lee, S. E. & Woodward, F. I. Contrasting physiological and structural vegetation feedbacks in climate change simulations. *Nature* **387**, 796–799 (1997).
- Enting, I., Wigley, T. & Heimann, M. *Future Emissions and Concentrations of Carbon Dioxide; Key Ocean/Atmosphere/Land Analyses* (Technical Paper 31, Division of Atmospheric Research, CSIRO, Melbourne, 1994).
- Gordon, C. *et al.* The simulation of SST, sea ice extents and ocean heat transports in a version of the Hadley Centre coupled model without flux adjustments. *Clim. Dyn.* **16**, 147–168 (2000).
- Palmer, J. R. & Totterdell, I. J. Production and export in a global ocean ecosystem model. *Deep-Sea Res.* (in press).
- Wilson, M. F. & Henderson-Sellers, A. A global archive of land cover and soils data for use in general circulation climate models. *J. Clim.* **5**, 119–143 (1985).
- Zinke, P. J., Stangenberger, A. G., Post, W. M., Emanuel, W. R. & Olson, J. S. *Worldwide Organic Soil Carbon and Nitrogen Data* (NDP-018, Carbon Dioxide Information Center, Oak Ridge National Laboratory, Oak Ridge, Tennessee, 1986).
- Cramer, W. *et al.* Global response of terrestrial ecosystem structure and function to CO₂ and climate change: results from six dynamic global vegetation models. *Glob. Change Biol.* (in press).
- VEMAP Members. Vegetation/ecosystem modelling and analysis project: comparing biogeography and biogeochemistry models in a continental-scale study of terrestrial responses to climate change and CO₂ doubling. *Glob. Biogeochem. Cycles* **9**, 407–437 (1995).
- Longhurst, A., Sathyendranath, S., Platt, T. & Caverhill, C. An estimate of global primary production in the ocean from satellite radiometer data. *J. Plank. Res.* **17**, 1245–1271 (1995).
- Field, C., Behrenfeld, M., Randerson, J. & Falkowski, P. Primary production of the biosphere: integrating terrestrial and oceanic components. *Science* **281**, 237–240 (1998).
- Antoine, D., Andre, J.-M. & Morel, A. Oceanic primary production 2. Estimation at global scale from satellite (Coastal Zone Color Scanner) chlorophyll. *Glob. Biogeochem. Cycles* **10**, 57–69 (1996).
- Tian, H. *et al.* Effects of interannual climate variability on carbon storage in Amazonian ecosystems. *Nature* **396**, 664–667 (1998).
- Keeling, C. D., Whorf, T., Whalen, M. & der Plicht, J. V. Interannual extremes in the rate of rise of atmospheric carbon dioxide since 1980. *Nature* **375**, 666–670 (1995).
- Houghton, J. T., Callander, B. A. & Varney, S. K. (eds) *Climate Change 1992: The Supplementary Report to the IPCC Scientific Assessment* (Cambridge Univ. Press, Cambridge, 1992).

- Nicholls, N. *et al.* in *Climate Change 1995: The Science of Climate Change* Ch. 3 (eds Houghton, J. T. *et al.*) (Cambridge Univ. Press, Cambridge, 1996).
- Mitchell, J. F. B., Johns, T. C., Gregory, J. M. & Tett, S. F. B. Climate response to increasing levels of greenhouse gases and sulphate aerosols. *Nature* **376**, 501–504 (1995).
- Wood, R. A., Keen, A. B., Mitchell, J. F. B. & Gregory, J. M. Changing spatial structure of the thermohaline circulation in response to atmospheric CO₂ forcing in a climate model. *Nature* **399**, 572–575 (1999).
- Sarmiento, J. & Quere, C. L. Oceanic carbon dioxide uptake in a model of century-scale global warming. *Nature* **274**, 1346–1350 (1996).
- Giardina, C. & Ryan, M. Evidence that decomposition rates of organic carbon in mineral soil do not vary with temperature. *Nature* **404**, 858–861 (2000).
- Orr, J. C. in *Ocean Storage of Carbon Dioxide, Workshop 3: International Links and Concerns* (ed. Ormerod, W.) 33–52 (IEA R&D Programme, CRE Group Ltd, Cheltenham, UK, 1996).
- Cox, P. M., Huntingford, C. & Harding, R. J. A canopy conductance and photosynthesis model for use in a GCM land surface scheme. *J. Hydrol.* **212–213**, 79–94 (1998).
- Cox, P. M. *et al.* The impact of new land surface physics on the GCM simulation of climate and climate sensitivity. *Clim. Dyn.* **15**, 183–203 (1999).
- Collatz, G. J., Ball, J. T., Grivet, C. & Berry, J. A. Physiological and environmental regulation of stomatal conductance, photosynthesis and transpiration: a model that includes a laminar boundary layer. *Agric. Forest Meteorol.* **54**, 107–136 (1991).
- Collatz, G. J., Ribas-Carbo, M. & Berry, J. A. A coupled photosynthesis-stomatal conductance model for leaves of C₄ plants. *Aust. J. Plant Physiol.* **19**, 519–538 (1992).
- Raich, J. & Schlesinger, W. The global carbon dioxide flux in soil respiration and its relationship to vegetation and climate. *Tellus B* **44**, 81–99 (1992).
- McGuire, A. *et al.* Interactions between carbon and nitrogen dynamics in estimating net primary productivity for potential vegetation in North America. *Glob. Biogeochem. Cycles* **6**, 101–124 (1992).

Acknowledgements

We thank J. Mitchell and G. Jenkins for comments on earlier versions of the manuscript. This work was supported by the UK Department of the Environment, Transport and the Regions.

Correspondence and requests for materials should be addressed to P.M.C. (e-mail: pmcox@meto.gov.uk).

Offset of the potential carbon sink from boreal forestation by decreases in surface albedo

Richard A. Betts

Hadley Centre for Climate Prediction and Research, The Met Office, Bracknell, Berkshire RG12 2SY, UK

Carbon uptake by forestation is one method proposed¹ to reduce net carbon dioxide emissions to the atmosphere and so limit the radiative forcing of climate change². But the overall impact of forestation on climate will also depend on other effects associated with the creation of new forests. In particular, the albedo of a forested landscape is generally lower than that of cultivated land, especially when snow is lying^{3–9}, and decreasing albedo exerts a positive radiative forcing on climate. Here I simulate the radiative forcings associated with changes in surface albedo as a result of forestation in temperate and boreal forest areas, and translate these forcings into equivalent changes in local carbon stock for comparison with estimated carbon sequestration potentials^{10–12}. I suggest that in many boreal forest areas, the positive forcing induced by decreases in albedo can offset the negative forcing that is expected from carbon sequestration. Some high-latitude forestation activities may therefore increase climate change, rather than mitigating it as intended.

Perturbations to the balance between radiation absorbed and emitted by the Earth ('radiative forcing') can result from changes in atmospheric chemistry and planetary albedo. A positive 'greenhouse' forcing results from increased atmospheric CO₂ absorbing and re-emitting more of the infrared radiation emitted by the surface¹³; forestation may help to mitigate this by slowing the rise

in atmospheric CO₂ concentrations through carbon sequestration¹², but may also provide a further positive forcing by reducing the surface albedo. Forests are generally darker than open land, particularly when snow is lying, because trees generally remain exposed whereas cultivated land can become entirely snow-covered^{3–5}. Snow-free foliage is considerably darker than snow, but even if large quantities of snow are held on the canopy, multiple reflections within the canopy scatter rather than reflect shortwave radiation which also reduces the landscape albedo⁴. Models suggest that the resulting low surface albedo causes boreal and cool-temperate forests to exert a warming influence on climate relative to unfor-ested land^{6–9} that may be more important than the effects of their carbon storage⁸.

To compare the forcings by albedo change and carbon sequestration due to mid- and high-latitude forestation, I have made spatially explicit calculations of shortwave radiation budget perturbations, considering replacement of agricultural land with forest. I then compare these with the effects of carbon sequestration by calculating the changes in atmospheric CO₂ concentration (and hence changes in carbon stock) that would give the same forcings as those exerted by the local albedo changes. For consistency with carbon sequestration estimates¹¹, I have considered dense coniferous plantations. These generally have a higher sequestration potential than natural forests at middle and high latitudes.

The radiative forcing due to surface albedo change was simulated with the radiative transfer scheme¹⁴ of the third Hadley Centre Atmosphere Model (HadAM3) (ref. 15). This simulates the Earth's radiation budget with a 3-hour timestep on a global grid of 3.75° longitude by 2.5° latitude, with 19 levels in the vertical. The short-wave radiation budget at each gridpoint is influenced by the simulated surface albedo and also by cloud and water vapour provided by other components of the atmosphere model. The surface albedo depends on the local vegetation and soil types and the depth and temperature of any lying snow¹⁶; of particular importance here is the inclusion of the effect of vegetation on surface albedo in snowy conditions⁶. The surface albedo α is calculated at each gridpoint as

$$\alpha = \alpha_0 + (\alpha_D - \alpha_0)(1 - e^{-0.25S}) \quad (1)$$

where α_0 is the local albedo in snow-free conditions, S is the snow mass (kg m⁻²), and α_D represents the influence of vegetation type and surface temperature $T^*(K)$ when snow is lying¹⁶:

$$\alpha_D = \begin{cases} \alpha_s & \text{for } T^* < T_M - \Delta T \\ \alpha_s + 0.3(\alpha_0 - \alpha_s)(T^* - T_M + \Delta T)/\Delta T & \text{for } T_M - \Delta T \leq T^* \leq T_M \\ \alpha_0 & \text{for } T^* > T_M \end{cases} \quad (2)$$

T_M is 273.15 K, and ΔT is 2.0 K. The vegetation-dependent deep-snow albedo parameter α_s provides the upper limit on albedo during deep snow cover, and is much lower for forests than for open land (Table 1). In snowy conditions, α can vary from 0.8 to 0.25 according to whether the local vegetation is farmland or forest.

Two 20-year radiation budget simulations were performed, with different data fields of α_0 and α_s but with the same meteorological

inputs. Simulation CONIF used α_0 and α_s for dense, mature, evergreen coniferous forest (Table 1) at points north of 23°N where widespread forest is sustainable^{17,18}, and simulation CROP used α_0 and α_s for arable cropland (Table 1) at these same points. The simulations calculated the surface albedo for the given vegetation cover using the same patterns of S and T^* from a 20-year HadAM3 simulation of present-day climate, and used this to model the shortwave radiation budget using cloud cover and other data again from HadAM3. The simulated outgoing shortwave flux at the tropopause was compared at each gridpoint, and the difference (CONIF–CROP) gave the local shortwave radiative forcing due to a change from cropland to coniferous forest.

In the boreal forest regions, the simulated surface albedo was 0.1–0.3 lower in CONIF than in CROP in the annual mean (Fig. 1a). The difference was greatest in the areas of longest-lasting snow cover, which were generally the farthest north. In temperate regions, the difference ranged from 0.04 to 0.15 depending on snow-cover duration and local soil albedo. The associated annual-mean local shortwave radiative forcing ranged from 3 W m⁻² in temperate regions to over 20 W m⁻² in the boreal forests of eastern Canada and eastern Siberia (Fig. 1b). The spatial maxima of the annual mean albedo and forcing did not coincide, because higher latitudes are sunlit for a shorter time and hence allow less forcing per unit albedo change.

To compare the climatic influence of albedo change with that of carbon sequestration for a given location, the effect of local albedo change on the global mean radiative forcing was found. The change in carbon stock that would give this same global forcing via atmospheric CO₂ change was then calculated. Because the calculations are nonlinear, the results will depend slightly on the scale of the forestation units; here, 1-ha plantations were considered. The contribution F of a plantation to global forcing was found by dividing the local radiative forcing (Fig. 1b) by the Earth's surface area in hectares, and the global-mean atmospheric CO₂ concentration change ΔC which would give the same forcing F was given by¹⁹

$$F = 5.35 \ln(1 + \Delta C/C_0) \quad (3)$$

where C_0 is the 1997 CO₂ concentration, 363.8 p.p.m.v. (ref. 20). ΔC is related to the terrestrial carbon stock change ΔC_T by

$$\Delta C_T = 2(M_c/M_a)m_a\Delta C/C_0 \quad (4)$$

where M_c and M_a are the molecular masses of carbon and dry air, and m_a is the mass of the atmosphere. The factor of 2 accounts for an airborne fraction of emissions of 0.5 (ref. 13), assumed to remain constant over forest growth timescales. Combining equations (1) and

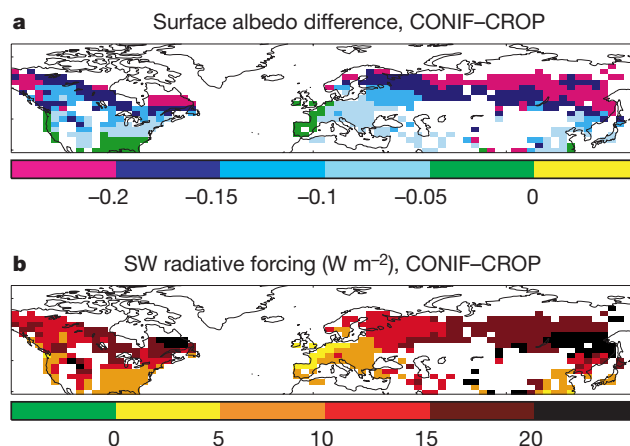


Figure 1 Effects of forestation on the solar radiation budget. **a**, Difference in annual-mean surface albedo α simulated by CONIF and CROP. **b**, Local instantaneous shortwave (SW) radiative forcing at the tropopause due to surface albedo change. At uncoloured gridpoints, vegetation was identical in CONIF and CROP.

Table 1 Albedo parameter values before and after forestation

Land cover	α_0 (dark soil)	α_0 (medium soil)	α_0 (light soil)	α_s
Arable cropland	0.18	0.19	0.21	0.78
Dense coniferous forest	0.14	0.14	0.15	0.26

Values of snow-free albedo parameter α_0 and deep-snow albedo parameter α_s for arable cropland and dense evergreen coniferous forest. α_0 depends on the albedo of the underlying soil, which varies according to soil type. Soil moisture dependence of α_0 is assumed to be negligible, and dependencies on wavelength and solar zenith angle are also neglected. These values are used along with those for other land surface types to derive α_0 and α_s fields for HadAM3 and the current operational weather forecast models of the Met Office^{2,29,30}.

(2) to find ΔC_T from F gave the emissions equivalent of the shortwave forcing (EESF). Assuming that α_0 and α_s reach the values given in Table 1 within one forest management rotation period, the EESF can then be compared with sequestration estimates^{10–12} (Table 2).

In the boreal forest regions, the replacement of crops with carbon-sequestering forests induced a positive albedo forcing equivalent to the emission of 50–140 t C ha⁻¹ (Fig. 2a). The greatest effect was in eastern Siberia, where the EESF was 90–140 t C ha⁻¹. The carbon sequestration potential (SP) in the former Soviet Union is estimated to be 80–120 t C ha⁻¹ (Table 2), so if this is appropriate to eastern Siberia then forestation in this region could exert a net warming influence (Fig. 2b). In the intensive agricultural regions of western Russia, the EESF of 80–90 t C ha⁻¹ could again largely counteract or overcome the estimated carbon sink (Fig. 2). Similarly in Canada, the EESF was 60–110 t C ha⁻¹, which significantly outweighs the mean sink potential of 60 t C ha⁻¹ estimated for most of this region (Fig. 2, Table 2). This suggests that in many boreal forest areas, forestation could therefore exert a net positive radiative forcing of climate, rather than a negative forcing as intended.

The EESF in British Columbia was smaller than the estimated SP (Table 2) because the relatively mild climate leads to less snow cover than elsewhere in Canada and also, presumably, enhances sequestration. However, the albedo change still significantly reduces the net forcing of climate in comparison with that expected from sequestration alone, because the net equivalent stock change (NESC = SP – EESF) is only 60% of the SP (Table 2). Similarly, the NESC for Nordic Europe is only 50% of the SP (Table 2). Carbon accounting¹² would therefore overestimate the associated mitigation of climate change by a factor of approximately two.

In temperate regions, the EESF ranged from 20 to 80 t C ha⁻¹ (Fig. 2a). The smallest effects were in western Europe, the western USA and southern China, where infrequent snow cover and/or low soil albedo¹⁷ suppressed the albedo change when the vegetation was modified. The effects were largest where snow lies for a significant part of the year; the EESF reached 80 t C ha⁻¹ in the northern USA, and exceeded 100 t C ha⁻¹ in the Rocky Mountains, northern China and the fringes of the Tibetan plateau. The area-mean EESF was smaller than the estimated SP in all regions classified here as ‘temperate’, and for most regions the NESC was 70–90% of the SP (Table 2). The mean NESC for China was only 30% of the SP, the latter being relatively small due to a low growth rate estimate¹¹ while the mean EESF was high because of large forcings in the north (Fig. 2a). I note that the growth rate given for this region is similar to that for the neighbouring former Soviet Union¹¹, so may be applicable only to northern China. Latitudinal comparisons with other regions suggest that higher SPs may be more appropriate in

southern China, but higher-resolution sequestration estimates will be required to confirm this. Despite such uncertainties, these results show that although temperate forestation should still exert a net negative radiative forcing, the effect cannot be adequately quantified by simple carbon accounting.

These results clearly rely on accurate representation of S , α_0 and α_s . S from HadAM3 agrees with an observation-based monthly climatology²¹ to within 10% across most regions in the annual mean. HadAM3 gives more snow than the climatology in parts of Russia, and less in eastern Europe, but the greatest disagreement is in southeastern Canada where the climatology gives around twice as much snow mass as HadAM3. The observation-based snow mass would reduce the EESF by approximately 10–15 t C ha⁻¹ in eastern Siberia, but would increase that in eastern Canada by around 15 t C ha⁻¹. Climate and weather simulations using α_0 and α_s as in Table 1 show no evidence of bias attributable to errors in these parameter values for these surface types. However, lower α_s values of

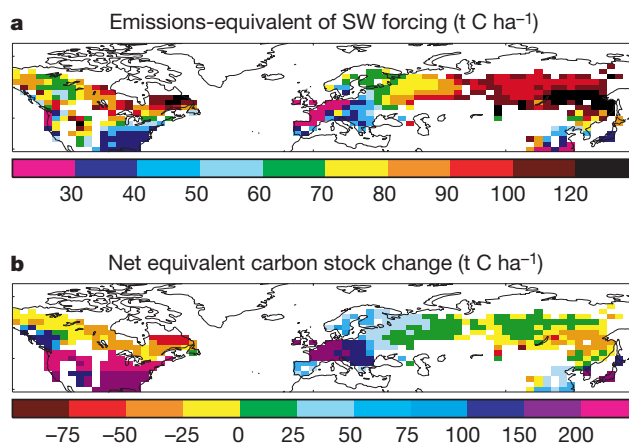


Figure 2 Effects of albedo change on annual-mean global radiative forcing in terms of carbon stock change. **a**, Carbon emissions (t C ha⁻¹) that would provide global radiative forcing equal to that exerted by local surface albedo change due to forestation (EESF). See text for error estimates. **b**, Net changes in equivalent carbon stock (NESC, t C ha⁻¹) considering both sequestration potential (SP) and EESF. SP contribution is from regionally representative values (Table 2) applied across whole regions because higher-resolution data are not available; this results in discontinuities across some region boundaries (for example, the USA/Canada border) where smooth transitions are expected in reality. Where ranges of SP estimates apply, range means are used here. Negative NESC implies a net radiative forcing equivalent to net carbon emission. At uncoloured gridpoints, vegetation was identical in CONIF and CROP.

Table 2 Sequestration and albedo forcings in terms of carbon stock change

Region	Sequestration potential, SP (t C ha ⁻¹)	Emissions-equivalent of shortwave forcing, EESF (t C ha ⁻¹)	Net equivalent stock change, NESC (t C ha ⁻¹)	NESC/SP (%)
<i>Boreal</i>				
Former Soviet Union	80–120	100	–20–20	–30–20
British Columbia	190	80	110	60
Rest of Canada	60	90	–30	–50
Nordic Europe	120	60	60	50
<i>Temperate</i>				
Western Europe	140–280	30	110–250	80–90
Eastern Europe	150	40	110	70
Southern Europe	90	40	50	60
Temperate USA	200–420	70	130–350	70–80
Southern USA	210	40	170	80
China	80	60	20	30
Rest of temperate Asia	200	60	140	70

Column 2 gives ranges of regional estimates of carbon sequestration potential by forestation over one rotation period, based on literature data^{10,11}. Coniferous plantations are specified for sequestration estimates in the boreal forest regions, the USA and western Europe^{10,11}; for other regions the forest type is unspecified¹¹ and was assumed here to be coniferous. Sequestration potentials include above- and below-ground biomass, and were estimated here as mean net uptake^{10,11} times mean rotation period^{10,11}. Rotation periods range from 40 to 80 yr. Nabuurs and Mohren¹⁰ give total system uptake at specified locations representative of major biomes. Nilsson and Schopfhauser¹¹ give regional values and separate above-ground and below-ground uptake, the latter being in root biomass (20%/19% of total living biomass for boreal/temperate forests¹¹), litter (0.32%/2.11% of above-ground biomass¹¹), and soil (regionally specific estimates¹¹). In column 4, NESC = SP – EESF. Data are rounded to the nearest 10 to avoid misleading precision.

0.10–0.18 have been suggested for dense coniferous forest^{4,5}; this would increase the EESF by approximately 25 tC ha⁻¹ in the most snow-covered regions.

Here I have considered forestation under present-day conditions, but the effects of future CO₂ rise and climate change are likely to affect the magnitude of both radiative forcing terms, due to dependencies on time-varying quantities such as the atmospheric CO₂ concentration, snow extent and vegetation structure and leafiness. As the atmospheric CO₂ concentration increases, CO₂ fertilization is likely to increase carbon uptake²² so the magnitude of the negative sequestration forcing should therefore increase, although associated climate changes may exert additional positive or negative effects on sequestration. Warmer temperatures may reduce the extent of snow cover²³, but the leaf area index (LAI) of potential vegetation may increase^{24,25}, so the albedo forcing could either increase or decrease. The effect of vegetation on surface albedo is not necessarily proportional to biomass, so the net contribution to radiative forcing may not evolve linearly throughout a forest's development; albedo depends on canopy density and architecture, and can become low rapidly, whereas carbon sequestration depends largely on woody biomass which is more gradually accumulated. Other contributions to forcing may also require consideration; for example, the longwave radiation budget could be affected by modified surface emissivity²⁵, although the sign of such changes is uncertain^{25,26}.

The work I report here has focused on perturbations to the Earth's radiation budget, which is the fundamental driver of the climate system. Forestation may also influence the climate by modifying the fluxes of heat, moisture and momentum between the land surface and atmosphere. Whereas boreal forests warm their local climate through reduced albedo, tropical forests tend to cool and moisten their local climates by greatly enhancing evaporation. Both may also influence distant regional climates via the atmospheric circulation^{9,27}. Assessment of the effect of forestation on climate at a given time in the future will require simulations with a climate model that incorporates vegetation dynamics^{25,28} and other atmospheric, terrestrial and oceanic components of the carbon cycle²⁸, in which forest growth occurs at appropriate rates in relation to changes in atmospheric CO₂ and snow cover. Nevertheless, my results suggest that high-latitude forestation would exert a positive radiative forcing through reduced albedo that in many places could outweigh the negative forcing through carbon sequestration. If afforestation and reforestation are required to decrease radiative forcing rather than simply to reduce net CO₂ emissions, then changes in surface albedo must also be considered. □

Received 10 July; accepted 27 September 2000.

1. UNFCCC Kyoto Protocol to the United Nations Framework Convention on Climate Change Art. 3.3 (UNEP/INC/98/2, Information Unit for Conventions, UNEP, Geneva, 1998) <http://www.unfccc.int/resource/docs/convkp/kpeng.pdf>.
2. UNFCCC United Nations Framework Convention on Climate Change Art. 2 (UNEP/IUC/99/2, Information Unit for Conventions, UNEP, Geneva, 1999); <http://www.unfccc.int/resource/convkp.html>.
3. Robinson, D. A. & Kukla, G. Albedo of a dissipating snow cover. *J. Climatol. Appl. Meteorol.* **23**, 1626–1634 (1984).
4. Harding, R. J. & Pomeroy, J. W. The energy balance of the winter boreal landscape. *J. Clim.* **9**, 2778–2787 (1996).
5. Sharratt, B. S. Radiative exchange, near-surface temperature and soil water of forest and cropland in interior Alaska. *Agric. Forest Meteorol.* **89**, 269–280 (1998).
6. Thomas, G. & Rowntree, P. R. The boreal forests and climate. *Q. J. R. Meteorol. Soc.* **118**, 469–497 (1992).
7. Bonan, G. B., Pollard, D. & Thompson, S. L. Effects of boreal forest vegetation on global climate. *Nature* **359**, 716–718 (1992).
8. Bonan, G. B., Chapin, F. S. & Thompson, S. L. Boreal forest and tundra ecosystems as components of the climate system. *Clim. Change* **29**, 145–167 (1995).
9. Douville, H. & Royer, J. F. Influence of the temperate and boreal forests on the Northern Hemisphere climate in the Météo-France climate model. *Clim. Dyn.* **13**, 57–74 (1997).
10. Nabuurs, G. J. & Mohren, G. M. J. Modelling analysis of potential carbon sequestration in selected forest types. *Can. J. Forest Res.* **25**, 1157–1172 (1995).
11. Nilsson, S. & Schophauser, W. The carbon sequestration potential of a global reforestation program. *Clim. Change* **30**, 267–293 (1995).
12. Watson, R. T. et al. (eds) *Land Use, Land-use Change and Forestry* (Cambridge Univ. Press, Cambridge, 2000).

13. Schimel, D. et al. in *Climate Change 1995. The Science of Climate Change* Ch. 2 (eds Houghton, J. T. et al.) 65–131 (Cambridge Univ. Press, Cambridge, 1995).
14. Edwards, J. M. & Slingo, A. Studies with a flexible new radiation code. I: Choosing a configuration for a large-scale model. *Q. J. R. Meteorol. Soc.* **122**, 689–720 (1996).
15. Pope, V. D., Gallani, M. L., Rowntree, P. R. & Stratton, R. A. The impact of new physical parametrizations in the Hadley Centre climate model - HadAM3. *Clim. Dyn.* **16**, 123–146 (2000).
16. Hansen, J. E. et al. Efficient three dimensional global models for climate studies, Models I and II. *Mon. Weath. Rev.* **111**, 609–662 (1983).
17. Wilson, M. F. & Henderson-Sellers, A. A global archive of land cover and soils data for use in general circulation climate models. *J. Climatol.* **5**, 119–143 (1985).
18. Woodward, F. I., Smith, T. M. & Emanuel, W. R. A global land primary productivity and phytogeography model. *Glob. Biogeochem. Cycles* **9**, 471–490 (1995).
19. Myhre, G., Highwood, E. J., Shine, K. P. & Stordal, F. New estimates of radiative forcing due to well mixed greenhouse gases. *Geophys. Res. Lett.* **25**, 2715–2718 (1998).
20. Keeling, C. D. & Whorf, T. P. *Atmospheric CO₂ Concentrations - Mauna Loa Observatory, Hawaii, 1958–1997* (NDP-001, Carbon Dioxide Information Analysis Centre, Oak Ridge, Tennessee, 1998).
21. Willmott, C. J., Rowe, C. M. & Mintz, Y. Climatology of the terrestrial seasonal water cycle. *J. Climatol.* **5**, 589–606 (1985).
22. Cao, M. & Woodward, F. I. Dynamic responses of terrestrial ecosystem carbon cycling to global climate change. *Nature* **393**, 249–252 (1998).
23. Essery, R. Seasonal snow cover and climate change in the Hadley Centre GCM. *Ann. Glaciol.* **25**, 362–366 (1997).
24. Betts, R. A., Cox, P. M., Lee, S. E. & Woodward, F. I. Contrasting physiological and structural vegetation feedbacks in climate change simulations. *Nature* **387**, 796–799 (1997).
25. Levis, S., Foley, J. A. & Pollard, D. Potential high-latitude vegetation feedbacks on CO₂-induced climate change. *Geophys. Res. Lett.* **26**, 747–750 (1999).
26. Kondratyev, K. Y., Korzov, V. I., Mukhenberg, V. V. & Dyachenko, L. N. in *Land Surface Processes in Atmospheric General Circulation Models* (ed. Eagleson, P. S.) 463–514 (Cambridge Univ. Press, Cambridge, 1982).
27. Gedney, N. & Valdes, P. J. The effect of Amazonian deforestation on the northern hemisphere circulation and climate. *Geophys. Res. Lett.* **27**, 3053–3056 (2000).
28. Cox, P. M., Betts, R. A., Jones, C. D., Spall, S. A. & Totterdell, I. J. Acceleration of global warming due to carbon-cycle feedbacks in a coupled climate model. *Nature* **408**, 184–187 (2000).
29. Cox, P. M. et al. The impact of new land surface physics on the GCM simulation of climate and climate sensitivity. *Clim. Dyn.* **15**, 183–203 (1999).
30. Posey, J. W. & Clapp, P. F. Global distribution of normal surface albedo. *Geophys. Int.* **4**, 333–348 (1964).

Acknowledgements

I thank S.E. Lee and F.I. Woodward for providing data from the Sheffield University vegetation model, and P.M. Cox, J.M. Edwards, R.L.H. Essery, W.J. Ingram, G.J. Jenkins, J.E. Lovelock, S. Nilsson, I.C. Prentice, P.R. Rowntree, K.P. Shine, P.J. Valdes and D.A. Warrilow for advice, comments and discussion. This work forms part of the Climate Prediction Programme of the UK Department of the Environment, Transport and the Regions.

Correspondence should be addressed to the author (e-mail: rabetts@meto.gov.uk).

An artificial landscape-scale fishery in the Bolivian Amazon

Clark L. Erickson

Department of Anthropology, University of Pennsylvania, 33rd and Spruce Streets, Philadelphia, Pennsylvania 19104-6398, USA

Historical ecologists working in the Neotropics argue that the present natural environment is an historical product of human intentionality and ingenuity, a creation that is imposed, built, managed and maintained by the collective multigenerational knowledge and experience of Native Americans^{1,2}. In the past 12,000 years, indigenous peoples transformed the environment, creating what we now recognize as the rich ecological mosaic of the Neotropics^{3–6}. The prehispanic savanna peoples of the Bolivian Amazon built an anthropogenic landscape through the construction of raised fields, large settlement mounds, and earthen causeways^{7,8}. I have studied a complex artificial network of hydraulic earthworks covering 525 km² in the Baures region of Bolivia. Here I identify a particular form of earthwork, the zigzag structure, as a fish weir, on the basis of form, orientation, location, association with other hydraulic works and ethnographic analogy.

The native peoples used this technology to harvest sufficient animal protein to sustain large and dense populations in a savanna environment.

The zigzag structure is a particular form of artificial earthwork, found in the seasonally inundated savanna of Baures, Bolivia (Province of Iténez, Department of the Beni) (Figs 1 and 2). Zigzag structures are linear segments of raised earth (1–2 m wide and 20–50 cm tall) that change direction every 10–30 m (Fig. 3). Shrubs, palms, and termite mounds cover the structures. Many zigzag structures cross the savanna from one forest island to another, distances of up to 3.5 km; others terminate 5,000–1,000 m from the forest edge (Fig. 4). Funnel-like openings, 1–3 m long and 1–2 m wide, are present where the structures form a sharp angle (Fig. 5). The structures are associated with small circular ponds. Dense networks of interconnected zigzag structures form enclosures of 10–80 ha. A total of 48,431 linear kilometres of weirs were measured in a sample area of 16,755 km² of savanna, a density of 2.891 linear km per km².

The zigzag structures are artificial constructions created by raising earth removed from the adjacent savanna. Canals or barrow pits flank some of the zigzag structures. Although overlapping in distribution, the zigzag structures are distinct from the long, wide and straight causeways and canals that cross the savanna between forest islands (Fig. 4). The narrow and irregular zigzag structures would be inefficient for transportation. The zigzag structures do not appear to have functioned as check dams or berms for flood-recessional farming. There is no evidence of crop furrows or field platforms between the structures.

On the basis of location, form, patterning, associations and ethnographic analogy, I identify the zigzag structures as fish weirs. Fish migrate to and spawn in the seasonally inundated savannas of Baures during the wet season⁹. Many fish are trapped in water bodies as the floodwaters recede. The zigzag structures provided a means to manage and harvest these fish. The zigzag structures are similar to fish weirs built by native peoples in Bolivia^{10–14} and throughout the Americas^{15–17}. Two characteristics shared by fish weirs include construction of barriers across shallow bodies of water and V-shaped openings where fish are trapped. Weirs, ranging in length from several to hundreds of metres, are constructed of soil, rock, reed, branches, logs, aquatic vegetation and/or basketry. Superstructures of perishable materials or a dense wall of vegetation probably covered the earthen fish weirs of Baures.

There are important differences between the zigzag structures of Baures and contemporary fish weirs. Most ethnographic fish weirs are ephemeral and rebuilt each season. Traditional weirs are constructed in rivers, streams or permanent bodies of water. In contrast, the zigzag structures are permanent earthworks built across a

seasonally flooded savanna. They are also more numerous, longer and more densely placed than ethnographic fish weirs. In addition to controlling and harvesting fish within the savanna, the weirs and large causeways may have been used for water management. The earthworks could have extended the period of inundation by capturing the first rains and holding floodwaters into the dry season¹⁸.

The savanna fisheries of the Bolivian Amazon are productive. Estimates of 100,000 to 400,000 fish have been calculated for a single hectare of abandoned river channel in the savannas⁹. Yields of 1,000 kg per hectare per year have been recorded for shallow ponds in tropical savannas¹⁹. Large numbers of *Pomacea gigas* snails are found beside the zigzag structures. In addition to fish, these edible snails may have been managed and raised in the weir structures and ponds. In the past, these snails were eaten in Baures and the gastropods are found in precolumbian sites in Bolivia and Brazil^{20,21}. The nutritional status of *Pomacea* is probably similar to other tropical snails, low in calories and protein²². *Pomacea gigas* reproduce and grow at an impressive rate and an average of 23.8 snails per m³ is recorded in Bolivian wetlands²³. The artificial fisheries of Baures potentially produced hundreds of tonnes of edible snails as a secondary food source.

The most common vegetation associated with the fish weirs and ponds is the palm *Mauritia flexuosa*^{24–26} (Fig. 3). A single tree can produce up to 5,000 edible fruits each year and a single hectare yields 10–60 t of fruit. The fruits are high in vitamins C and A, oil (12%) and protein (4–5% dry weight). The ground tissue produces large amounts of edible starch. Edible larvae of the palm beetle thrive in the decomposing trunks. In addition, the palm is a favoured food of game animals and fish. The fibres of the fronds and trunks are used for basketry, mats, hammocks, bowstrings, thatch and roof beams. The palm may have been encouraged or even cultivated on the earthworks.

Artificial ponds overlap with the distribution of zigzag structures (Fig. 3). These measure 0.5–2 m deep and 10–30 m in diameter; the largest hold water year round. The ponds teem with fish such as buchere (*Hoplosternum* sp.), yallu, cunaré (*Cichla monoculus*), palometa (*Serrasalmus* sp.), sábalo (*Prochilodus nigricans*) and bentón (*Erythrina* sp), snails, birds, reptiles and amphibians. Contemporary hunters stalk the game animals and birds that congregate at the ponds. Artificial ponds provided a way to store live fish and snails until needed.

The complex of fish weirs and ponds of Baures is a form of intensive aquaculture. The earthworks did not necessarily involve the mobilization of large amounts of labour. I estimate a total of 1,515 linear kilometres of fish weirs in Baures based on a sample of aerial photographs. Using labour estimates for experimental

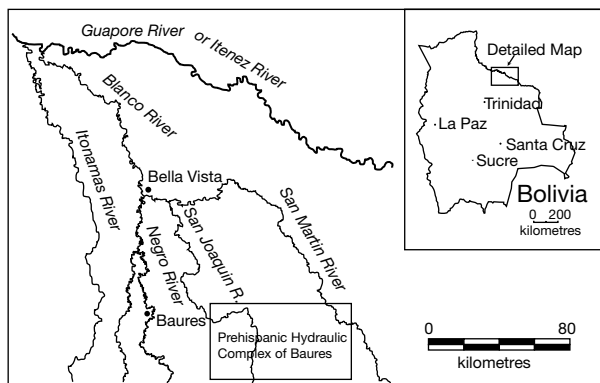


Figure 1 Map of the Baures prehispanic hydraulic complex. It is located between the San Joaquín river to the west and the San Martín river to the east and between 13° 30' and 14° 20' latitude South.

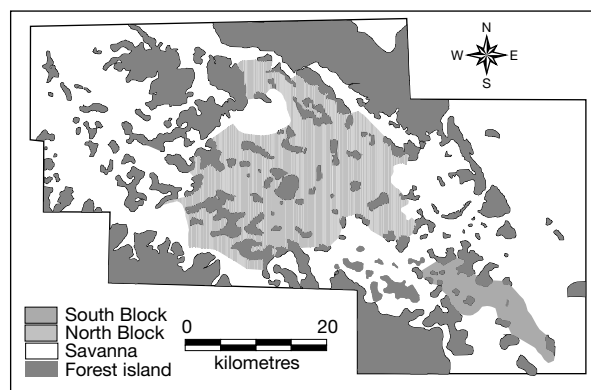


Figure 2 The fish weirs (zigzag structures and ponds). They are concentrated in the North Block (447 km² of savanna excluding forest islands) and the South Block (77 km²). The structures are restricted to savannas near forest islands that are flooded by shallow water.

construction of raised fields (5 m^3 of earth per person per 5-hr day and weirs measuring 2 m wide and 0.5 m tall), the weirs required 300,000 person-days of labour or the equivalent of 1,000 people working 30 days a year over a period of 10 years. Small groups of kin or communities constructed and managed the weirs recorded in ethnographic accounts^{10–14}. Similar social groups may have been

responsible for the weirs and ponds of Baures.

The weirs also show some evidence of integration at a higher scale. Individual zigzag structures often cross the savanna from one forest island to another (Fig. 4). Assuming each forest island was an autonomous settlement, weir construction may have involved inter-community cooperation. Although individual weirs could operate



Figure 3 Oblique photograph of a fish weir and artificial ponds between forest islands in the savannas of Baures. Fish weirs are the zigzag structures, lower left to upper right;

artificial ponds are the circular features surrounded by palms (approximately 20 m in diameter). The diagonal feature (upper left to lower right) is a contemporary path.

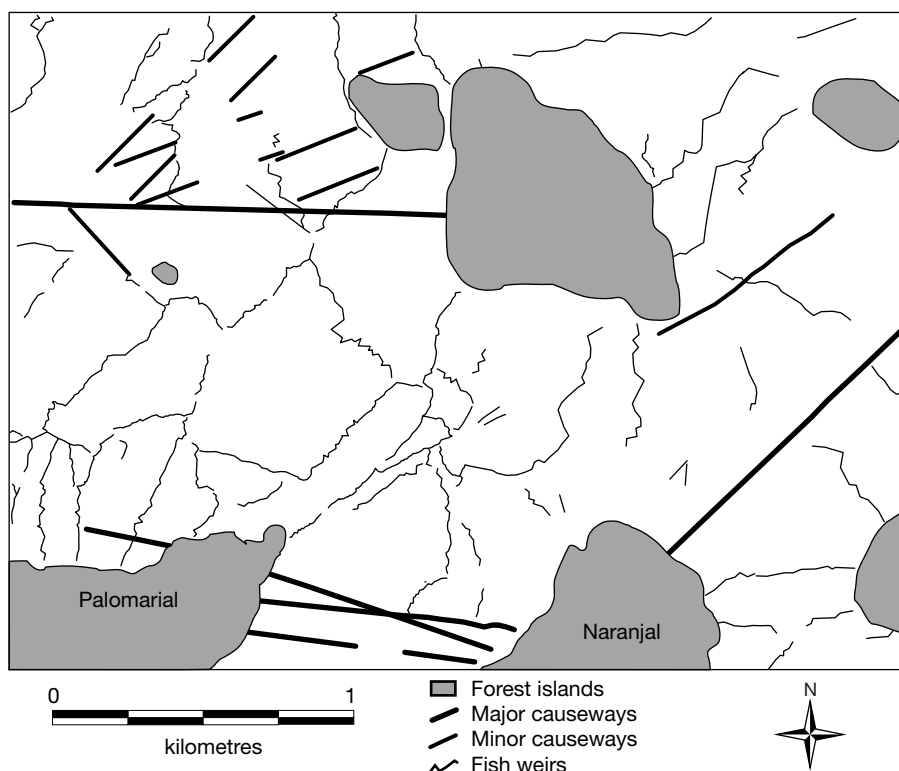


Figure 4 Map of fish weirs (irregular lines) and causeways (straight lines) in Baures. Based on aerial photographs.

independently of other weirs, the presence of an integrated network of causeways, canals, weirs and ponds suggest a higher order of water management^{8,18}. As a permanent food-producing infrastructure, the weirs must have been valuable real estate. The networks of causeways and canals may have promoted communication and alliances between individual communities exploiting the fish weirs. Groups in the Colombian Amazon jealously protect and guard riverine fisheries, valuable resources that are owned and inherited by clans and chiefly lineages¹⁶. The presence of moated, and presumably palisaded, settlements on many of the forest islands suggests potential tension over the fisheries and other resources²⁷.

Colonial accounts describe the use of causeways in Baures for communication and transportation between settlements^{7,10,12,27}. Weirs in lakes and streams are described^{10,12}, but there is no mention of the zigzag structures in the savanna. To date the fish weirs, I excavated a large causeway directly associated with zigzag structures²⁸. Burned wood from the base of the causeway fill was radiocarbon dated to 335 years BP (before present) $\pm$ 20 (OS-17293) or an uncalibrated calendar date of AD 1615 (AD 1595–1635). The corrected date at 68.2% confidence is AD 1490 (0.26) AD 1530; AD 1560 (0.74) AD 1630. Depending on the context, the sample may date or predate the original construction. The Spanish did not control the Baures region until 1708; thus, the earthwork probably predates European occupation.

The earthworks of Baures are an example of creation and active management of an anthropogenic landscape by native peoples. The linear causeways and canals were a sophisticated means of regulating water levels within the savannas to enhance and manage seasonal aquatic resources. The network of fish weirs provided a means of controlling and harvesting fish, in addition to enhancing the habitat and availability of aquatic and terrestrial fauna. The artificial ponds were a means of concentrating and storing live fish, providing drinking water and improving game habitats. Palms growing on earthworks provided additional foodstuffs and materi-

als. Using this simple, but elegant, technology, the people of Baure converted much of the landscape into an aquatic farm covering 500 km². Rather than domesticate the species that they exploited, the people of Baure domesticated the landscape. The fish weirs and ponds produced abundant, storable, and possibly sustainable yields of animal protein. Thus, they were able to sustain large dense populations in what many would consider a marginal environment. □

Received 2 July; accepted 1 September 2000.

1. Posey, D. & Balée, W. (eds) *Resource Management in Amazonia: Indigenous and Folk Strategies* (Advances in Economic Botany Vol. 7, New York Botanical Gardens, Bronx, 1989).
2. Balée, W. (ed.) *Advances in Historical Ecology* (Columbia Univ. Press, New York, 1998).
3. Stahl, P. W. Holocene biodiversity: An archaeological perspective from the Americas. *Annu. Rev. Anthropol.* **25**, 105–126 (1996).
4. Denevan, W. M. The pristine myth: The landscape of the Americas in 1492. *Ann. Ass. Am. Geog.* **82**, 369–385 (1992).
5. Piperno, D. & Pearsall, D. *Origins of Agriculture in the Lowland Neotropics* (Academic, San Diego, 1998).
6. Roosevelt, A. C. et al. Paleoindian cave dwellers in the Amazon: The peopling of the Americas. *Science* **272**, 373–384 (1996).
7. Denevan, W. M. *The Aboriginal Cultural Geography of the Llanos de Mojos of Bolivia* (Univ. California Press, Berkeley, 1966).
8. Erickson, C. L. in *Archaeology in the American Tropics: Current Analytical Methods and Applications* (ed. Stahl, P.) 66–95 (Cambridge Univ. Press, Cambridge, 1995).
9. Hanagarth, W. *Acerca de la geoeología de las sabanas del Beni en el noreste de Bolivia* 134 (Instituto de Ecología, La Paz, 1993).
10. Eder, F. J. *Breve Descripción de las Reducciones de Mojos* (Historia Boliviana, Cochabamba, 1985 [1772]).
11. Jones, J. *Conflict between Whites and Indians on the Llanos de Mojos, Beni Department Dissertation*. Univ. Florida (1980).
12. de Castillo, J. in *Documentos para la Historia Geográfica de la República de Bolivia: Las Provincias de Mojos y Chiquitos* (ed. Ballivián, M. V.) Vol. 1 308–317 (J. M. Gamarrá, La Paz, 1906 [1676]).
13. Riestler, J. *Los Guarasugwe: Crónica de sus últimos días* (Editorial Los Amigos del Libro, La Paz, 1977).
14. Ortiz Lema, E. *Los Matacos Nocturnos de Bolivia* (Editorial los Amigos del Libro, La Paz, 1986).
15. Beckerman, S. in *Amazonian Indians from Prehistory to the Present: Anthropological Perspectives* (ed. Roosevelt, A.) 177–200 (Univ. Arizona Press, Tucson, 1994).
16. Chernela, J. M. *The Wanano Indians of the Brazilian Amazon: A Sense of Place* (Univ. Texas Press, Austin, 1993).
17. Nordenskiöld, E. *The Ethnography of South America seen from Mojos in Bolivia* Comparative Ethnographical Studies 3 86–102 (Göteborg, 1924).
18. Erickson, C. L. in *Vías Precolombinas: Los Caminos, Los Ingenieros Y Los Viajeros* (eds Herrera, L. & Cardale de Schimpff, M.) (Editorial Abya Yala y El Instituto Colombiano de Antropología, Bogotá, Colombia, in the press).
19. Garson, A. Comment upon the economic potential of fish utilization in riverine environments and the potential archaeological biases. *Am. Antiquity* **45**, 62–567 (1980).
20. Dougherty, B. & Calandra, H. *Relaciones de la Sociedad Argentina de Antropología* Vol. 16, 37–61 *Ambiente y arqueología en el oriente boliviano: La Provincia de Iténez del Departamento Beni*. (1984–5).
21. Schmitz, P. I., Rogge, J., Rosa, A. & Beber, M. *Aterros indígenas no Pantanal do Mato Grosso do Sul* (Pesquisas Antropologia no. 54, Instituto Anchieta de Pesquisas, Rio Grande do Sul, 1998).
22. Moholy-Nagy, H. The utilization of Pomacea snails at Tikal, Guatemala. *Am. Antiquity* **43**, 65–73 (1978).
23. Swing, C. K., Isley, W. & Lutz, M. W. Notas sobre la ecología reproductiva del caracol. *Pomacea gigas*. *Ecología en Bolivia* **10**, 19–31 (1987).
24. Clement, C. R. in *Tropical Forests, People and Food* (eds Hladik, C. M. et al.) 139–154 (The Parthenon, Paris, 1993).
25. Beckerman, S. The abundance of protein in Amazonia: A reply to Gross. *Am. Anthropol.* **81**, 533–561 (1989).
26. Smith, N. J. H. *The Amazon River Forest: A Natural History of Plants, Animals and People* (Oxford Univ. Press, New York, 1999).
27. Lee, K. *Complejo Hidráulico de las llanuras de Baures (Area a ser protegida)*, Provincia Iténez, Departamento del Beni, República de Bolivia (unpublished report). (CORDEBENI, Trinidad, 1995).
28. Erickson, C. L., Winkler, W. & Candler, K. *Las investigaciones arqueológicas en la región de Baures en 1996* (unpublished report). (Instituto Nacional de Arqueología de Bolivia, La Paz, 1997).

Acknowledgements

Fieldwork and analysis were supported by the NSF, the Heinz Charitable Trust Foundation, the American Philosophical Society, the Research Funds of the University of Pennsylvania Museum, and Corporación del Beni. K. Lee originally reported the earthworks of Baures in the late 1950s. A. Vranich, O. Saavedra and F. Bruckner did the first study of the Baures earthworks in 1995. I thank the Bolivian Dirección Nacional de Arqueología y Antropología, the Prefectura and Alcaldía of the Department of the Beni, authorities of Baures and Bella Vista, W. Winkler (project co-investigator), K. Lee, H. Schlink, R. Bottega, R. Pinto Parada, E. Bruckner, A. Bruckner, C. Bruckner, O. Rivera, R. Langstroth, W. Denevan, A. Vranich, P. Stahl, R. Dunn and D. Brinkmeier.

Correspondence and requests for materials should be addressed to C.E. (e-mail: cerickso@sas.upenn.edu).

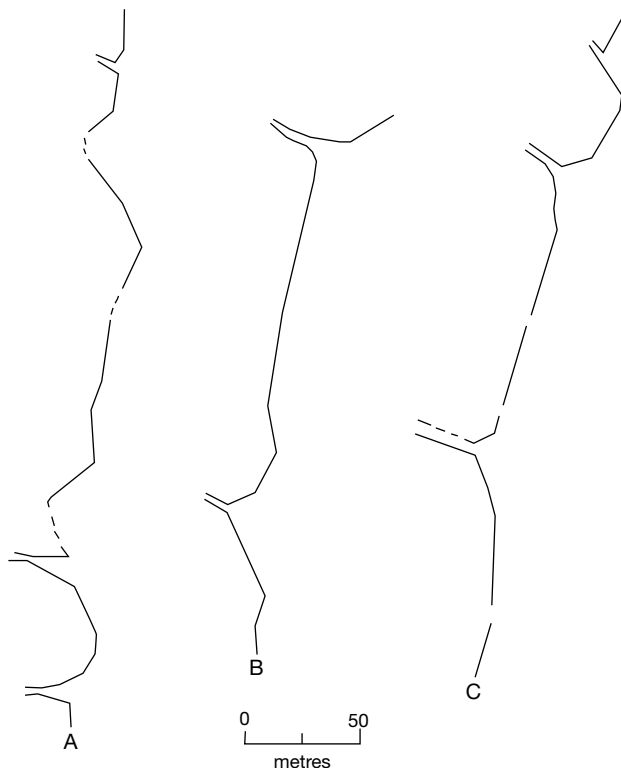


Figure 5 Plans of fish weirs (zigzag structures). Small parallel openings in the weirs are present every 50–200 m.

Spatial synchronization of vole population dynamics by predatory birds

Rolf A. Ims* & Harry P. Andreassen

Department of Biology, Division of Zoology, University of Oslo, PO Box 1050 Blindern, N-0316 Oslo, Norway

* Present address: NINA, Polar Environment Centre, N-9005 Tromsø, Norway

Northern vole populations exhibit large-scale, spatially synchronous population dynamics^{1,2}. Such cases of population synchrony provide excellent opportunities for distinguishing between local intrinsic and regional extrinsic mechanisms of population regulation³. Analyses of large-scale survey data and theoretical modelling^{4–6} have indicated several plausible synchronizing mechanisms. It is difficult, however, to determine the most important one without detailed data on local demographic processes^{3,7}. Here we combine results from two field studies in southeastern Norway—one identifies local demographic mechanisms and landscape-level annual synchrony among 28 enclosed experimental populations and the other examines region-level multi-annual synchrony in open natural populations. Despite fences eliminating predatory mammals and vole dispersal, the growth rates of the experimental populations were synchronized and moreover, perfectly linked with vole abundance in the region. The fates of 481 radio-marked voles showed that bird predation was the synchronizing mechanism. A higher frequency of risky dispersal movements in slowly growing populations appeared to accelerate predation rate. Thus, dispersal may induce a feedback-loop between predation and population growth that enhances synchrony.

Theoretically, three mechanisms may induce population synchrony³: regionalized climatic disturbances; dispersal between populations; and predators focusing on prey hotspots. Survey data^{1,8,9} suggest that the spatial extent of synchrony in boreal voles (about 30–50 km) is smaller than would be expected from the action of climate. Thus, dispersal and predation are more likely mechanisms. However, distinguishing between them based on survey data alone has proved to be difficult^{9,10}; it seems that experiments controlling for dispersal and/or predation are needed. Moreover, different vole predators, especially birds and

mammals, need to be distinguished because they operate on different temporal and spatial scales¹¹.

The two studies were carried out in boreal southeastern Norway where vole populations are known to fluctuate considerably over time and show pronounced inter- and intraspecific spatial synchrony². The experimental study took place at Evenstad in the Glomma valley (61° 12' N, 11° 06' E). The landscape at Evenstad consists of boreal forest interrupted by patches of agricultural fields and meadows. We used the root vole *Microtus oeconomus* as the experimental species. It is a herbivorous small rodent (body mass, 40–90 g), temporally common in small meadow patches at Evenstad. The first week of July (week 26) of each of the four years 1990–1993, we established standardized root vole populations on 7 separate 50 × 100 m fence-enclosed meadow plots. Electric wire-mesh fences hindered intrusion of mammalian predators and steel-sheet fences hindered dispersal of voles. Avian predators were allowed to enter.

The experimental populations were monitored closely by weekly live-trapping for three and a half months, corresponding to the length of the breeding season of northern root voles. There was a large year-to-year variation in the development of the populations over the breeding season (Fig. 1). Estimates of year-specific mean per capita growth rate (r) and their 95% confidence limits were all positive in 1990–1992 (1990: $r = 0.64$ (0.28, 0.99), 1991: $r = 0.96$ (0.80, 1.13), 1992: $r = 0.83$ (0.39, 1.26)), while the populations declined over the breeding season in 1993 ($r = -0.53$ (–0.95, –0.11)). The yearly effect accounted for 58% of the total variance in population growth rate among the 28 population replicates and thus there was evidence ($\chi^2 = 29.04$, degrees of freedom (d.f.) = 3, $P < 0.0001$) for a within-year landscape-level synchronization.

Predatory birds specialized on voles, especially migrating owls (*Asio otus* and *Asio flammeus*) and raptors (*Falco tinnunculus* and *Buteo lagopus*) were seen hunting on the experimental plots. However, because owls hunt at night we did not attempt to monitor the activity of these predators directly. Instead we quantified predation rates by equipping reproductive female voles ($n = 481$) with radio-transmitters. Predation-induced mortality, known from the fates of radio-marked females, was the only demographic parameter accounting for a significant portion of the variability in population growth (predation effect, $\chi^2 = 29.44$, d.f. = 1, $P < 0.0001$). The proportion of voles caught by birds explained 61% of the total variation and 75% of the year-to-year variation in population growth rate. Predation was exceptionally high in 1993 (chance of predation in 1993 was 5.31 (95% confidence interval (c.i.) = 2.95, 9.02) times higher than in 1990–1992) causing population declines

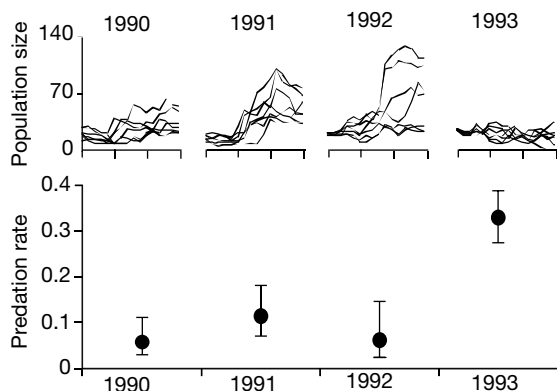


Figure 1 Population development and predation rate of the experimental populations. Upper panel, population trajectories (number of individuals per population) of the 28 experimental populations of root voles. Lower panel, mean predation rate per week attributable to avian predators in the four years. The predation rate estimates are based on radio-tracking of adult females in the populations during one week in August and one week in September. Error bars are 95% confidence intervals based on a logistic-binomial model.

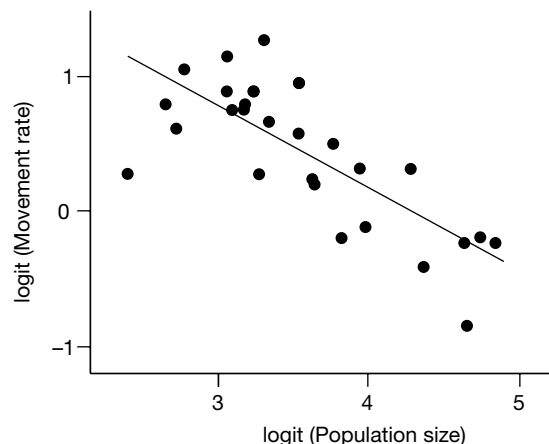


Figure 2 Negative density-dependent movement rates. The relationship between proportion of animals that had moved at least once beyond patch borders (as determined from weekly live trapping) and the size of the 28 population replicates in week 38. Week 38 corresponds with peak population size in most of the populations (see Fig. 1). Logistic regression: $\chi^2 = 54.9$, d.f. = 1, $P < 0.0001$.

over the summer (Fig. 1). Fecundity did not differ between years (analysis of variance (ANOVA): $F = 1.7$, d.f. = 1, 26, $P = 0.3814$).

Small mammals are most vulnerable to predation when moving in marginal habitats^{12,13}. The experimental plots were managed so that habitat patches were imbedded in open matrix areas with little cover^{14,15}. Movements outside habitat patches increase predation by birds^{12,15}. Moreover, the frequency of such dispersal movements increased with declining density (Fig. 2); that is, negative growth rate-dependent or density-dependent dispersal appears to be common in mammals^{16,17}. Taken together, this suggests a feedback loop between predation and population growth mediated by dispersal.

The large-scale transect survey, specifically designed to determine the spatial scale of small rodent population synchrony in this geographic region, was sampled over the years 1990–1994 (ref. 8). If avian predation (the mechanism evidently causing synchrony among the experimental populations) also synchronized vole dynamics at a regional scale, we expected the growth rates of the experimental populations to correlate positively with regional vole abundance. That is, in a region of low prey density, local hotspots (such as our experimental populations) should be subject to spatially density-dependent predation¹⁸ and, consequently, low growth rate. Indeed, there was almost perfect correspondence (r_p (Pearson correlation coefficient) = 0.98, $n = 4$, $P = 0.0167$) between the average growth rate in the experimental populations and the corresponding average abundance index of voles over the years 1990–1993 (Fig. 3). Including data from the final year (1994) of the transect survey and an additional experimental year at Evenstad (Fig. 3) strengthened the statistical significance of this relationship ($r_p = 0.99$, $n = 5$, $P = 0.0017$).

By combining experimental and observational data we have shown that vole population dynamics was tightly synchronized at two spatial and temporal scales; within-year synchrony at landscape scale (synchrony between the experimental populations) and between-year synchrony at the regional scale (matching growth

rates of the experimental populations and regional abundance level). The very controlled setting of the experimental populations, with standardized initial conditions every year and with fences to hinder vole dispersal and intrusion by predatory mammals, allowed us to eliminate all potential intrinsic and most extrinsic determinants of population synchrony. Variable summer climate has to our knowledge never been proposed as a determinant of northern vole population dynamics^{19,20}, although it could potentially affect diet quality. However, fecundity and body growth (the two demographic parameters most sensitive to summer diet²⁰) did not exhibit significant year-to-year variation. On the other hand, predation-induced mortality was the main demographic variance component in the experimental populations and the only probable synchronizing mechanism. In fact, models¹⁸ predict that summer predation rates much lower than we observed are sufficient to cause population synchrony.

The influence of predatory birds was strong enough to induce a decline in the experimental populations over the breeding season in 1993. Such summer declines characterize northern vole populations^{19,20} and are believed to be caused mainly by small mustelids, which include weasels and stoats^{21–24}. Here we have shown that predatory birds alone can induce summer declines, at least in vole hotspots when there is a regionally low abundance level of voles. Birds that particularly hunt voles have the capacity for rapid and precise prey-tracking²⁵ and sometimes inflict predation rates in natural vole populations^{26–28} as high as those in our experimental populations.

Our study suggests an unusual negative feedback loop between predation and prey population growth, mediated by risky dispersal movements, which acts to enhance the synchronizing effect of predation. The causes of growth rate and/or density-dependent dispersal at the individual level and their consequences at the population and the meta-population level deserve further study. □

Methods

Experimental populations

Seven experimental populations per year were initiated by releasing 3–5 laboratory-raised mothers and their just-weaned litters into each enclosure (size, 50 m × 100 m). The initial population densities averaged 153 ± 10 (s.e.m.) individuals per hectare of meadow habitat which correspond to peak densities in natural root vole populations. Demography was monitored on a weekly basis with two days of trapping per week from early July (week 26) until mid-October (week 42) when all animals were removed so that the enclosures were empty until the next summer. In total 2,929 individuals were captured 32,055 times. Potential aberrant effects of frustrated movements beyond the enclosures were prohibited by removing animals that were frequently trapped in particular fence traps following the same rule for removal every year¹⁴. Removed animals did not make up more than 8.6% of the total number of individuals in any population, and the removals did not affect demography¹⁴. Except for three populations in 1992, all reproductive females in the experimental populations were equipped with radio-transmitters (SS20, Biotrack, UK) during one week in August and one week in September. Individuals found depredated or disappeared from the enclosures during these two weeks could reliably be assigned as victims of avian predation¹⁵. The radio-transmitters (collars) themselves do not impair survival²⁹.

The vegetation was burned and fertilized every spring before the introduction of voles to standardize habitat productivity. Population growth rate in a particular enclosure in year t was unrelated to population growth rate in that enclosure in year $t - 1$ ($\chi^2 = 0.02$, d.f. = 1, $P = 0.871$) indicating that habitat quality indeed became standardized by this procedure. The spatial habitat configuration within the seven plots was designed to create a mosaic of meadow patches imbedded in a barren matrix. The two main designs were temporally replicated in 1990 and 1991 (ref. 14) and in 1992 and 1993 (ref. 15), respectively. Although the size and number of the habitat patches varied between the designs, the amount of habitat became the same. In the log-linear population growth model, habitat configuration accounted for less than 13% of the variance ($\chi^2 = 8.55$, d.f. = 5, $P = 0.0734$). Other factors that could potentially contribute to the year-to-year variation in population growth rate, such as the number of female lineages (matrilines) initiating each population, their initial mean body weight and body growth rate in the field, were even less important ($P > 0.16$).

Demographic analyses of the experimental populations

We used a log-linear version of the general stochastic growth model $N_{\text{week}42} = \text{quasi-Poisson}[N_{\text{week}26} \times \exp(c + \Sigma a_i \times X_i)]$ to analyse the per capita population growth between onset (week 26) and termination of the experimental periods (week 42) with respect to the various predictor terms ($a_i \times X_i$). These were year, predation rate, habitat-patch configuration, the number of animals released, their mean body weight at release and body growth in the field. The growth model was fitted with logarithmic link function and with $N_{\text{week}26}$ as an offset term and quasi-Poisson error which included only

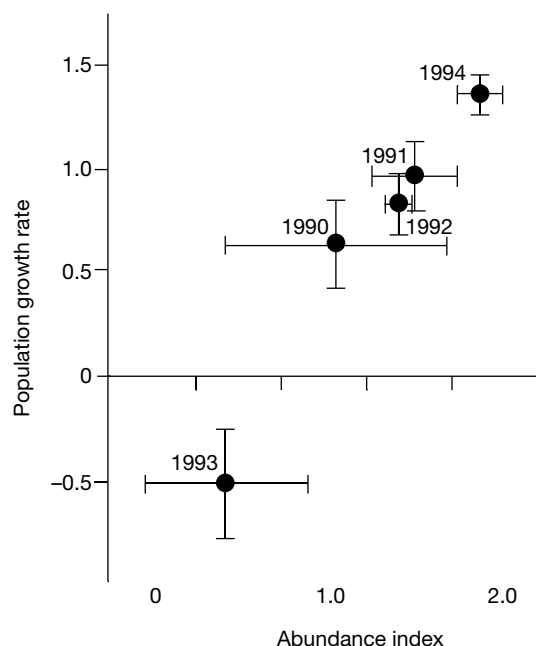


Figure 3 Evidence for regional multi-annual spatial synchrony. The relationship between yearly mean growth rates of experimental root vole populations at Evenstad and yearly mean abundance indices (log number of individuals per trapping site) of natural vole populations in the general area. The data from the experimental plots at Evenstad 1994 is based on six experimental root vole populations that were initiated in the same manner and in a similar setting as in the four preceding years (1990–1993). 1994 was not included in the main analysis of the demography, because the trapping protocol was less intensive¹² and there was no radio-tracking to determine predation rate.

process variance, as population sizes at the onset and termination of the experimental seasons could be determined without sampling error.

Predation rates were estimated by applying a logistic regression model on the population-specific proportion of radio-tracked females depredated by birds. Fecundity was estimated as the product of mean litter delivery rate and mean litter size of all adult females (older than 40 days) in each population.

Transect survey of regional vole abundance variation

We used trapping data from three trapping sites (numbers 19–21)⁸ located at the same altitude (250 m) in the neighbouring valley to Glomma, the Rena valley. The habitat and the landscape features were similar to those surrounding the experimental plots at Evenstad. The distances between Evenstad and the transect sites (19–34 km) were shorter than the local synchrony domain (about 40 km) of small rodent populations along the transect⁸. The mean number of microtine rodents (*Clethrionomys glareolus*, *Microtus agrestis* and *M. oeconomus* combined) snap-trapped per station each year was used as a yearly abundance index. Pooling the species is justified because of inter-specific population synchrony².

Received 2 August; accepted 21 August 2000.

1. Ranta, E. & Kaitala, V. Travelling waves in vole population dynamics. *Nature* **390**, 456 (1997).
2. Hansson, L. & Henttonen, H. Rodent dynamics as community processes. *Trends Ecol. Evol.* **3**, 195–200 (1988).
3. Bjørnstad, O. N., Ims, R. A. & Lambin, X. Spatial population dynamics: analysing patterns and processes of population synchrony. *Trends Ecol. Evol.* **14**, 427–432 (1999).
4. Ranta, E., Kaitala, V. & Lundberg, V. P. The spatial dimension in population dynamics. *Science* **278**, 1621–1623 (1997).
5. Grenfell, B. T. *et al.* Noise and determinism in synchronized sheep dynamics. *Nature* **394**, 674–677 (1998).
6. Blasius, B., Huppert, A. & Stone, L. Complex dynamics and phase synchronization in spatially extended ecological systems. *Nature* **399**, 354–359 (1999).
7. May, R. M. Crash tests for the real. *Nature* **398**, 371–372 (1999).
8. Steen, H., Ims, R. A. & Sørnerud, G. A. Spatial and temporal patterns of small rodent population dynamics at a regional spatial scale. *Ecology* **77**, 2365–2372 (1996).
9. Bjørnstad, O. N., Stenseth, N. C. & Saitoh, T. Synchrony and scaling in dynamics of voles and mice in northern Japan. *Ecology* **80**, 622–637 (1999).
10. Lambin, X. *et al.* Spatial asynchrony and periodic travelling wave in cyclic field vole populations. *Proc. R. Soc. Lond. B* **265**, 1491–1496 (1998).
11. Korpimäki, E. & Krebs, C. J. Predation and population cycles of small mammals. A reassessment of the predation hypothesis. *BioScience* **46**, 754–764 (1996).
12. Aars, J., Johannsen, E. & Ims, R. A. Demographic consequences of movements in subdivided root vole populations. *Oikos* **85**, 204–216 (1999).
13. Steen, H. Low survival of long distance dispersers of the root vole (*Microtus oeconomus*). *Ann. Zool. Fennici* **31**, 271–274 (1994).
14. Ims, R. A. & Andreassen, H. Effects of experimental habitat fragmentation and connectivity on root vole demography. *J. Anim. Ecol.* **68**, 839–852 (1999).
15. Andreassen, H. P. & Ims, R. A. The effect of experimental habitat destruction and patch isolation on space use and fitness parameters in female root vole *Microtus oeconomus*. *J. Anim. Ecol.* **67**, 941–952 (1998).
16. Clobert, J. *et al.* (eds) *Causes and Consequences of Dispersal at the Individual, Population and Community Level* (Oxford Univ. Press, in the press).
17. Wolff, J. O. Population regulation in mammals: an evolutionary perspective. *J. Anim. Ecol.* **66**, 1–13 (1997).
18. Ims, R. A. & Steen, H. Geographic synchrony in microtine population cycles: a theoretical evaluation of the role of nomadic avian predators. *Oikos* **57**, 381–387 (1990).
19. Krebs, C. J. Population cycles visited. *J. Mamm.* **77**, 8–24 (1996).
20. Stenseth, N. C. & Ims, R. A. *The Biology of Lemmings* (Academic, London, 1993).
21. Hanski, I. *et al.* Population oscillation of boreal rodents: regulation by mustelid predators leads to chaos. *Nature* **364**, 232–235 (1993).
22. Turchin, P. & Hanski, I. An empirically based model for latitudinal gradients in vole population dynamics. *Am. Nat.* **149**, 842–874 (1997).
23. Heikkilä, J., Below, A. & Hanski, I. Synchronous dynamics of microtine rodent populations on islands in Lake Inari in northern fennoscandia: evidence for regulation by mustelid predators. *Oikos* **70**, 245–252 (1994).
24. Korpimäki, E. & Norrdahl, K. Experimental reduction of predators reverses the crash phase phase of small-rodent cycles. *Ecology* **79**, 2448–2455 (1998).
25. Viitala, J. *et al.* Attraction of kestrels to vole scent marks visible in ultraviolet-light. *Nature* **373**, 425–427 (1995).
26. Steen, H. Untangling the causes of disappearance from a local population of root vole, *Microtus oeconomus*: a test of the regional synchrony hypothesis. *Oikos* **73**, 65–72 (1995).
27. Reid, D. G., Krebs, C. J. & Kenney, A. Limitation of collared lemming population-growth at low densities by predation mortality. *Oikos* **73**, 387–398 (1995).
28. Norrdahl, K. & Korpimäki, E. Mortality factors in a cyclic vole populations. *Proc. R. Soc. Lond. B* **261**, 49–53 (1995).
29. Johannsen, E., Andreassen, H. P. & Steen, H. C-J-S modelling revealed no effect of radio collars on survival of root voles. *J. Mamm.* **78**, 638–642 (1996).

Acknowledgements

We thank all the people who helped us during the field work at Evenstad; and J. Aars, E. Johannsen, X. Lambin and H. Steen for comments on the manuscript. This study was supported by the Research Council of Norway.

Correspondence and requests for materials should be addressed to R.A.I. (e-mail: r.a.ims@bio.uio.no).

Tracking an object through feature space

Erik Blaser*, Zenon W. Pylyshyn* & Alex O. Holcombe†

* Rutgers Center for Cognitive Science, Rutgers University, Piscataway, New Jersey 08854, USA

† Department of Psychology, Harvard University, Cambridge, Massachusetts 02138, USA

Visual attention allows an observer to select certain visual information for specialized processing. Selection is readily apparent in 'tracking' tasks where even with the eyes fixed, observers can track a target as it moves among identical distractor items¹. In such a case, a target is distinguished by its spatial trajectory. Here we show that one can keep track of a stationary item solely on the basis of its changing appearance—specified by its trajectory along colour, orientation, and spatial frequency dimensions—even when a distractor shares the same spatial location. This ability to track through feature space bears directly on competing theories of attention, that is, on whether attention can select locations in space^{2–4}, features such as colour or shape^{5–7}, or particular visual objects composed of constellations of visual features. Our results affirm, consistent with a growing body of psychophysical^{8–13} and neurophysiological^{14–16} evidence, that attention can indeed select specific visual objects. Furthermore, feature-space tracking extends the definition of visual object¹⁷ to include not only items with well defined spatio-temporal trajectories¹⁸, but also those with well defined featural-temporal trajectories.

In this investigation, observers tracked and made judgements about a circular striped 'Gabor' patch that dynamically changed its orientation, spatial frequency and colour, but never its spatial location. Along these feature dimensions the Gabor smoothly changed: spinning, say, clockwise for a while then counter-clockwise; changing gradually back and forth between a few broad stripes and many thin stripes; and altering its saturation, ranging smoothly between a patch made up of grey and black stripes to a patch made up of red and black stripes, through all the intervening saturation levels of red.

Our aim was to test whether observers can track such an item through feature space, but we were also interested in determining which of location-, feature-, or object-based theories of visual attention could best account for this ability, if it existed. Location and 'objecthood' can be difficult to distinguish, as typically a single object occupies a single location^{19,20}. We separated location from objecthood in our experiments by completely superimposing two such Gabor patches (Fig. 1a) such that there was no spatial distinction to support selection by location-based attention: individual parts of each Gabor changed dynamically, and were at the limits of the resolution of location-based attention^{4,21,22} (the spatial frequency of the 1-degree-diameter Gabor ranged dynamically between 1.5 to 8 cycles per degree; thus the width of a single stripe ranged from 20 to 4 arcmin). The Gabors appeared to observers as simultaneously present and transparently layered on one another, although there was no noticeable separation in depth (Fig. 1b).

In experiment 1, observers were instructed to track one of the two superimposed Gabors. During the tracking interval, the Gabors frequently 'passed' each other along one or more dimensions (Fig. 1c). (This ensured that feature-based attention could not simply pick out the target on the basis of some constant featural difference. Unless this is done, results that are consistent with object-based attention are subject to criticisms owing to the possible contributions of feature-based attention. For instance, this is of

concern in studies where two stimuli always have, say, different motion directions throughout the entire trial⁷. Although location-based attention may be foiled by the superimposition, feature-based attention may still pick out one of the stimuli by selectively enhancing its direction of motion.) The four observers (naive: JT, DA, DK; expert: EB) successfully picked the target Gabor with 90% accuracy.

How were observers able to accomplish this tracking through feature-space? Observers reported that the two superimposed

Gabors perceptually segregated from one another (although occasionally they appeared as a single entity, typically when the two objects had a near-miss or actually 'collided' in feature space). The segregation of superimposed stimuli has been observed in studies of motion transparency^{5,23}. In addition, judging from informal observations of our stimuli, the simultaneous changes along the orientation, colour and spatial frequency dimensions result in more compelling transparency than that which results from changes along any one, or pair, of these dimensions alone. Observers also reported that attending to a particular Gabor resulted in an increase of its salience, in a manner not unlike figure-ground segmentation. The attended Gabor became the figure and the distractor Gabor receded into the ground^{24,25}. As our stimuli were constructed to eliminate contributions from location- and feature-based attention, the ability to track through feature space alone is suggestive evidence for object-based attention.

Although they are not mutually exclusive, location-, feature- and object-based theories deal with our compound Gabor stimulus quite differently. Feature-based theories treat the stimulus as a mix of colour, orientation and spatial frequency information, any of which may be selectively enhanced by attention; location-based theories treat the stimulus as a unitary stream of information, the whole of which may be enhanced²⁶; but only object-based attention theories can treat the stimulus as two unitary streams of information, each corresponding to one of the superimposed Gabors, and each a legitimate target for visual attention. Moreover, object-based theories predict that when observers attempt to attend to a particular feature of one of the Gabors in our stimulus they actually, by default, attend to the object as a whole; consequently, processing is enhanced for all of its features¹⁰. This leads to two predictions that can be tested psychophysically. First, in contrast to location-based theories, object-based theories predict that selective enhancement of an object can take place even if there is another object that is spatially superimposed. Second, in contrast to feature-based theories, selective attention to any feature of an object should also enhance processing of its other features.

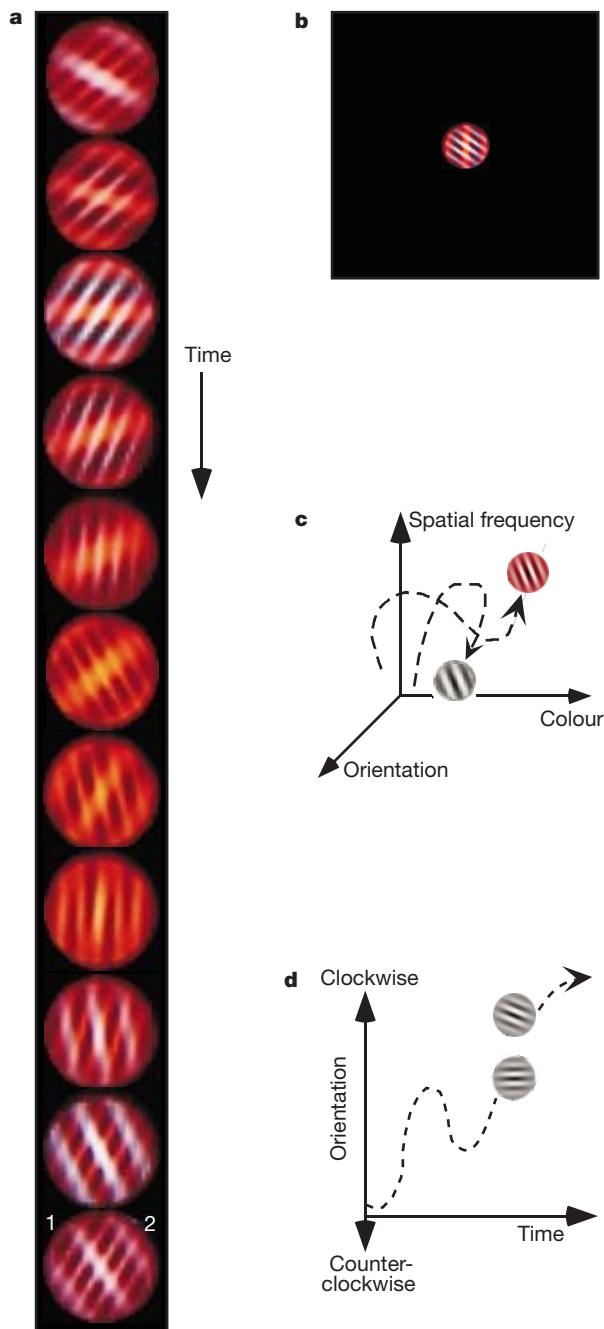


Figure 1 Objects in feature space. **a**, Snapshots taken every 250 ms of two superimposed Gabor patches drifting in a feature space defined by the dimensions of colour, orientation and spatial frequency. In experiment 1, observers tracked one target Gabor for 10 seconds, then picked out the target by reporting one of two small number labels (shown here in the last snapshot). **b**, Schematic of display, illustrating that the two Gabors were always stationary in space, and superimposed, at fixation. **c**, Diagram of the two Gabors drifting in feature-space. **d**, Diagram of a clockwise 'jump'. Gabors also jumped along the colour and frequency dimensions.

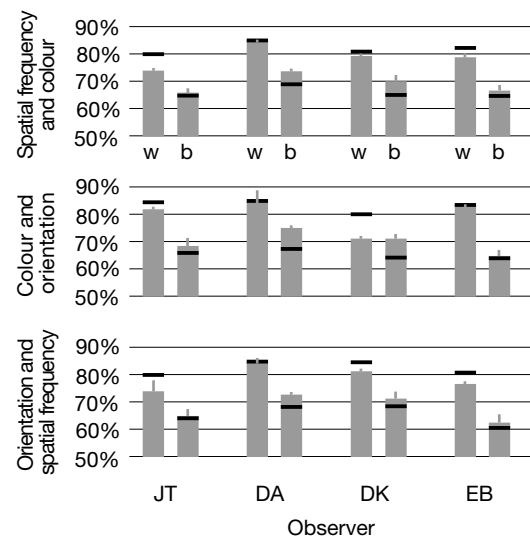


Figure 2 Experiment 2 results; percentage correct jump-direction judgements for each condition and observer. Conditions are labelled on the left, and specify the pair of feature dimensions observers were instructed to attend to, and make judgements about, for that particular block of trials. Pairs of bars show performance in within-object (w) and between-object (b) conditions, respectively. Spikes on the bars indicate standard errors. The black line above each within-object bar is the average performance from single-task control conditions, which is the theoretical maximum achievable performance. The black line on each between-object bar is the theoretical minimum performance. Data is shown for three naive (JT, DA, DK) observers and one expert observer (EB).

In experiment 2, we tested these predictions by instructing observers to perform two concurrent tasks. Performance in concurrent tasks reveals how the tasks compete for processing resources. When two tasks do not compete for resources, performance on one should not suffer when the other is performed concurrently (like walking and chewing gum). When two tasks do compete, performance on one of the tasks will suffer when the other task is performed concurrently (like juggling and sign language). Because competition occurs when an attention mechanism is forced to divide its resources, it is possible to determine which type of attention is mediating performance by observing when performance losses occur^{27,28}.

We introduced a small discontinuity, or “jump”, into the trajectory of each target along each dimension and then compared conditions for which a pair of jump-direction judgements was made within a common Gabor (for example, reporting the direction of the colour and orientation jumps for a specified Gabor), with conditions where the same two judgements were divided between the two Gabors (for example, reporting the direction of the colour jump for one Gabor, and the orientation jump for the other). If object-based attention mediates performance in this task, then processing will be enhanced for all of the features of one of the Gabors, but not the features of the other Gabor. Therefore, there should be little or no competition for resources provided that the concurrent judgements are made with respect to a common Gabor. However, if concurrent judgements are divided between the two Gabors, significant competition should result, as in that case attentional resources must somehow be divided between the objects.

This is exactly the pattern of results we found. When observers were asked to make two judgements about the same Gabor, performance was on average 80% correct, very close to the theoretical maximum performance of 83% correct given by single-task control conditions (when observers only had to make one judgement at a time). In other words, observers could make both of these judgements simultaneously almost as well as they made each of them alone (like walking and chewing gum). In contrast, when observers attempted to divide their attention to make the judgements between the two Gabors, performance dropped to 69% correct, quite close to the theoretical minimum of 66% correct that would be achieved if observers could not do both tasks at the same time at all (like juggling and sign language) and instead, for instance, did one task on the even trials and the other on the odd trials (Fig. 2). This asymmetry in the results cannot be accounted for by either location- or feature-based theories.

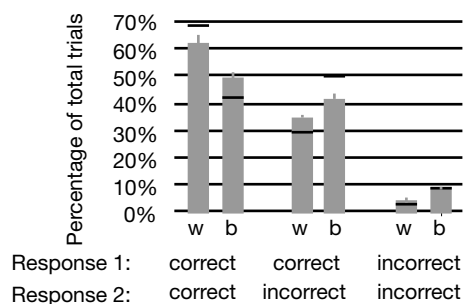


Figure 3 Assessment of ‘sharing’ versus ‘switching’ models of attention resources. Observers made two responses in both within-object (w) and between-object (b) conditions; thus, trials fall into three categories (correct/correct, correct/incorrect, or incorrect/incorrect). Bars correspond to the actual percentage of trials in each category, averaged over observer. Spikes on the bars indicate standard errors. Black lines on the within-object bars are the predictions of a pure sharing model, where it is assumed that observers perform both tasks at levels given by single-task control conditions. Black lines on the between-object bars are the predictions of a pure switching model, where observers perform one task at single-task levels, but perform the other at chance.

Observers are thus capable of tracking a single object in spite of a spatially superimposed distractor. But are observers able to track multiple objects simultaneously? In the between-object conditions of experiment 2, observers had both an instruction and a task that encouraged them to attend and track two objects simultaneously. It is clear that observers did much worse in these conditions than in the within-object conditions, where they had only to attend and track a single object. Nevertheless, perhaps observers were actually able to attend to and track both objects at the same time to some extent, that is, to ‘share’ attentional resources between the two. On the other hand, perhaps observers were unable to share and instead ‘switched’ attention—sometimes attending to one object, and sometimes attending to the other.

Though both sharing and switching models can account for overall performance, they make very different predictions for the underlying pattern of responses²⁷. Specifically, if observers were sharing, then the correctness of responses on the two judgements should show statistical independence. That is, whether or not observers were correct on the jump-direction judgement with respect to one object should not significantly influence their correctness with respect to the other. Alternatively, if observers were switching, then if observers were correct on one judgement, they should be less likely to be correct on the other, and vice versa. A χ^2 test of the between-object data rejected the null hypothesis of statistical independence ($P < 0.005$ for each observer), and therefore of pure sharing of attention. (As predicted by object-based theories of attention, this test performed on the within-object data cannot reject the hypothesis of statistical independence—indicating that observers share resources within an object, that is, simultaneously attend to all its features.) We then used the data from the single-task control conditions in conjunction with pure sharing and pure switching models to predict the proportion of trials in which observers should get both responses correct, both incorrect, and one correct and one incorrect. These predicted values were compared to the frequencies in the actual data (Fig. 3). The pattern is clear: observers approach pure sharing when attending to a single object (because resources were, by default, shared among all the object’s features), but approach pure switching when attempting to attend to two objects (because attention could only be devoted to a single object at a time). The result of this analysis, taken together with observers’ reports of an inability to attend to both objects simultaneously, indicates that object-based attention limits tracking through feature space at a given location to a single object.

The ability to track through feature-space and our finding of competition for attentional resources between, but not within, superimposed items, implies the operation of object-based attention, but also shows that distinct ‘visual objects’²⁹ need not have distinct spatiotemporal trajectories. Rather, distinct *feature*-temporal trajectories are sufficient for objecthood and attentional tracking. □

Methods

Stimuli

The basic stimulus was a circular striped ‘Gabor’ patch (windowed cosinusoid) that dynamically changed its orientation, spatial frequency and colour, but never its location on the screen. This can be conceptualized as a single item that smoothly drifted in a feature-space defined by orientation, spatial frequency and colour dimensions. The trajectory of the Gabor along each of the feature dimensions was random and independent, frequently changing direction and speed along any given dimension. The Gabor did, however, drift with some ‘inertia’ that made it more likely to continue drifting, along a particular dimension, in the direction it was going (the inertia factor simply places a probability on whether or not the object will change its speed and/or direction along a particular dimension at a given time). In all experiments and conditions, two such Gabors were completely spatially superimposed (Fig. 1a). The Gabors were superimposed by temporally interleaving their images at a rate of 117 Hz (58.5 Hz per item, with one Gabor shown in the even video frames and the other in the odd frames). Observers were seated 100 cm from the 1.76-cm stimulus, which subtended 1 degree of visual angle. The stimulus was presented on an otherwise dark screen. Observers were instructed to maintain fixation on the stimulus during presentation (Fig. 1b). As the trajectories of the Gabors were independent, they frequently ‘passed’ each other along one or more dimensions (Fig. 1c).

Tracking in feature-space

In experiment 1, observers were presented with the compound stimulus for an initial 1-second fixation interval, giving them an opportunity to attend to the target Gabor, which was identified by being initially oriented 45 degrees clockwise. Both Gabors then began changing along all three feature dimensions for a 10-second tracking interval. At the end of the tracking interval, the Gabors stopped changing, but remained on the screen. Two small number labels appeared above the stimulus (Fig. 1a). One of the labels was aligned with the orientation of one of the Gabors (for instance, if the Gabor ended up tilted, say, a bit clockwise, its label would appear at '1 o'clock'), whereas the other label was aligned with the other Gabor. Observers used a keypress to report the label that they thought corresponded to the target item. Observers received positive feedback and a point score to encourage best performance.

Object-based attention

As in experiment 1, in experiment 2, observers viewed the static compound stimulus for an initial 1-second fixation interval. Then both Gabors began changing along all three possible feature dimensions for a 5-s tracking interval. At some random time in the tracking interval, both Gabors exhibited a slight discontinuity in their featural trajectories simultaneously along all three dimensions (colour, spatial frequency and orientation); that is, there was a slight 'jump' in the trajectory of each Gabor (Fig. 1d). The directions of the jumps were chosen randomly and independently for each Gabor and dimension, and the sizes of the jumps were fixed at values corresponding to 75% correct jump-direction thresholds determined from baseline psychometric functions of jump-size for each dimension and observer. In any given block of trials, observers were instructed to attend concurrently to a pair of dimensions. After the tracking interval, both Gabors stopped changing, but remained on the screen. Observers then made a keypress to report the direction of the jump (that is, clockwise or counter-clockwise, more or less red, higher or lower spatial frequency) for the pair of dimensions. Single-task control conditions were also run, where observers only needed to attend to, and make jump-direction judgements about, one feature dimension. Observers received positive feedback and a point score.

Received 9 June; accepted 11 September 2000.

- Pylyshyn, Z. W. & Storm, R. W. Tracking multiple independent targets: Evidence for a parallel tracking mechanism. *Spatial Vision* **3**, 179–197 (1988).
- Posner, M. I. Orienting of attention. *Q. J. Exp. Psychol.* **32**, 3–25 (1980).
- Sagi, D. & Julesz, B. Enhanced detection in the aperture of focal attention during simple discrimination tasks. *Nature* **321**, 693–695 (1986).
- Eriksen, C. W. & Hoffman, J. E. Temporal and spatial characteristics of selective encoding from visual displays. *Percept. Psychophys.* **12**, 201–204 (1972).
- Cavanagh, P. Attention-based motion perception. *Science* **257**, 1563–1565 (1992).
- Wolfe, J. M., Cave, K. R. & Franzel, S. L. Guided search: an alternative to the feature integration model for visual search. *J. Exp. Psychol. Hum. Percept. Perform.* **15**, 419–433 (1989).
- Corbetta, M., Miezin, F. M., Dobmeyer, S., Shulman, G. L. & Petersen, S. E. Attentional modulation of neural processing of shape, color, and velocity in humans. *Science* **248**, 1556–1559 (1990).
- Duncan, J. Selective attention and the organization of visual information. *J. Exp. Psychol. Gen.* **113**, 501–517 (1984).
- Valdes-Sosa, M., Cobo, A. & Pinilla, T. Transparent motion and object-based attention. *Cognition* **66**, B13–B23 (1998).
- Luck, S. J. & Vogel, E. K. The capacity of visual working memory for features and conjunctions. *Nature* **390**, 279–281 (1997).
- Kahneman, D., Treisman, A. & Gibbs, B. J. The reviewing of object files: object-specific integration of information. *Cogn. Psychol.* **24**, 175–219 (1992).
- Baylis, G. C. & Driver, J. Visual attention and objects: evidence for hierarchical coding of location. *J. Exp. Psychol. Hum. Percept. Perform.* **19**, 451–470 (1993).
- Treisman, A. Feature binding, attention and object perception. *Philos. Trans. R. Soc. Lond. B* **353**, 1295–1306 (1998).
- Driver, J., Baylis, G. C. & Rafal, R. D. Preserved figure-ground segregation and symmetry perception in visual neglect. *Nature* **360**, 73–75 (1992).
- Valdes-Sosa, M., Bobes, M. A., Rodriguez, V. & Pinilla, T. Switching attention without shifting the spotlight object-based attentional modulation of brain potentials. *J. Cogn. Neurosci.* **10**, 137–151 (1998).
- O'Craven, K. M., Downing, P. E. & Kanwisher, N. fMRI evidence for objects as the units of attentional selection. *Nature* **401**, 584–587 (1999).
- Treisman, A. & Paterson, R. Emergent features, attention, and object perception. *J. Exp. Psychol. Hum. Percept. Perform.* **10**, 12–31 (1984).
- von Hofsten, C. & Spelke, E. S. Object perception and object-directed reaching in infancy. *J. Exp. Psychol. Gen.* **114**, 198–212 (1985).
- Lappin, J. S. Attention in the identification of stimuli in complex visual displays. *J. Exp. Psychol.* **75**, 321–328 (1967).
- Bonnel, A. M. & Prinzmetal, W. Dividing attention between the color and the shape of objects. *Percept. Psychophys.* **60**, 113–124 (1998).
- He, S., Cavanagh, P. & Intriligator, J. Attentional resolution and the locus of visual awareness. *Nature* **383**, 334–337 (1996).
- Blaser, E., Sperling, G. & Lu, Z. L. Measuring the amplification of attention. *Proc. Natl Acad. Sci. USA* **96**, 11681–11686 (1999).
- Qian, N., Andersen, R. A. & Adelson, E. H. Transparent motion perception as detection of unbalanced motion signals. I. Psychophysics. *J. Neurosci.* **14**, 7357–7366 (1994).
- Neisser, U. & Becklen, R. Selective looking: Attending to visually specified events. *Cogn. Psychol.* **7**, 480–494 (1975).
- Rock, I. & Gutman, D. The effect of inattention on form perception. *J. Exp. Psychol. Hum. Percept. Perform.* **7**, 275–285 (1981).
- Shih, S. I. & Sperling, G. Is there feature-based attentional selection in visual search? *J. Exp. Psychol. Hum. Percept. Perform.* **22**, 758–779 (1996).

- Sperling, G. & Melchner, M. J. The attention operating characteristic: examples from visual search. *Science* **202**, 315–318 (1978).
- Lee, D. K., Koch, C. & Braun, J. Attentional capacity is undifferentiated: concurrent discrimination of form, color, and motion. *Percept. Psychophys.* **61**, 1241–1255 (1999).
- Scholl, B. J. & Pylyshyn, Z. W. Tracking multiple items through occlusion: clues to visual objecthood. *Cognit. Psychol.* **38**, 259–290 (1999).

Acknowledgements

This study was supported by an NIH Grant to Z.P. and an NRSA Institutional Postdoctoral Fellowship to E.B. We thank P. Cavanagh for helpful discussion and comments.

Correspondence and requests for materials should be addressed to E.B. (e-mail: blaser@ruccs.rutgers.edu).

Distinct functions of the two isoforms of dopamine D2 receptors

Alessandro Usiello^{*†}, Ja-Hyun Baik^{*†},
Françoise Rougé-Pont[‡], Roberto Picetti^{*}, Andrée Dierich^{*},
Marianne LeMeur^{*}, Pier Vincenzo Piazza[‡] & Emiliana Borrelli^{*}

^{*}Institut de Génétique et de Biologie Moléculaire et Cellulaire, CNRS, INSERM, ULP, BP 163, 67404 Illkirch Cedex, C.U. de Strasbourg, France

[‡]INSERM U259, Université V. Segalen Bordeaux 2, Domaine de Carreire Rue C. Saint-Saens, 33077 Bordeaux Cedex, France

[†]These authors contributed equally to this work.

Signalling through dopamine D2 receptors governs physiological functions related to locomotion, hormone production and drug abuse^{1–7}. D2 receptors are also known targets of antipsychotic drugs that are used to treat neuropsychiatric disorders such as schizophrenia⁸. By a mechanism of alternative splicing, the D2 receptor gene encodes two molecularly distinct isoforms⁹, D2S and D2L, previously thought to have the same function. Here we show that these receptors have distinct functions *in vivo*; D2L acts mainly at postsynaptic sites and D2S serves presynaptic autoreceptor functions. The cataleptic effects of the widely used antipsychotic haloperidol¹ are absent in D2L-deficient mice. This suggests that D2L is targeted by haloperidol, with implications for treatment of neuropsychiatric disorders. The absence of D2L reveals that D2S inhibits D1 receptor-mediated functions, uncovering a circuit of signalling interference between dopamine receptors.

Dysfunctions of the dopaminergic system are involved in neurological disorders such as Parkinson's disease, Tourette's syndrome, schizophrenia and in pituitary tumours¹. Dopamine acts through membrane receptors of the seven transmembrane domain G-protein coupled family. Two classes of dopamine receptor have been defined: D1-like (D1R and D5R) and D2-like (D2R, D3R and D4R) which, respectively, stimulate and inhibit adenylyl cyclase, thereby regulating intracellular cAMP levels¹.

D2 receptors are highly expressed in the striatal complex and pituitary gland. Ablation of this receptor results in locomotor impairment^{2–4}, altered response to drug abuse⁵, pituitary tumours^{6,7} and the modification of the electrophysiological characteristics of D2R-expressing neurons^{10,11}. Thus, D2Rs have an essential position at the postsynaptic level, and, by acting as autoreceptors^{10,12}, in the regulation of the dopaminergic system by modulating dopamine release.

D2R has two isoforms, D2L and D2S, which are generated by alternative splicing⁹ and co-expressed in a ratio favouring the long isoform, D2L (Fig. 1c, wild type). D2L differs from D2S by the presence of an additional 29 amino acids within the third intracellular loop. This region is implicated in the receptor interaction

with G-proteins and indeed D2L and D2S possess different coupling affinities for these proteins^{13,14}. But until now, the roles of D2L and D2S have been considered equivalent.

To address whether these receptors act in cooperation or synergy, or whether their activity is completely independent and/or antagonistic, mice were generated in which expression of the D2L isoform has been specifically ablated. To selectively delete D2L receptor expression by homologous recombination, exon 6 of the D2R gene⁹ was replaced with a pGK-neomycin resistance cassette (Fig. 1a). The targeting vector was electroporated into D3 embryonic stem cells², and two clones isolated, of which one was injected into recipient blastocysts (Fig. 1b). Chimaeras were bred with C57BL/6 mice. D2L^{-/-} mice in a 75% BL/6 background were identified (Fig. 1b). Mutant mice appear healthy and fertile and indistinguishable from wild-type (WT) littermates.

RNAse protection assays performed on RNA isolated from WT and D2L^{-/-} tissues (Fig. 1c) revealed a protected fragment corresponding to D2L only in WT mice (Fig. 1c). D2S messenger RNA is upregulated in D2L^{-/-} mice (Fig. 1c), suggesting that, in the absence of exon 6 and flanking intronic regions, the post-transcriptional processing of the gene proceeds by default¹⁵. The insertion of the neo cassette in the genome of D2L^{-/-} mice did not alter the normal pattern of D2R expression in the brain (data not shown). Loss of D2L did not modify expression of other D2-like receptors, such as D3R (data not shown).

The pharmacological characteristics of D2S in D2L-null mice were assessed using ³H-spiperone (a D2 specific antagonist). In spite of the lack of D2L, a similar number of binding sites was found in D2L^{-/-} mice (maximum binding capacities as mean $\pm$ standard error of the mean, s.e.m., $B_{\max}$ = 273 $\pm$ 63 fmol mg⁻¹ protein) in

comparison to WT littermates (307 $\pm$ 46 fmol mg⁻¹ protein). The affinity for ³H-spiperone was unaltered (dissociation constant K_d as mean $\pm$ s.e.m.; K_d = 24 $\pm$ 1.5 pM for WT, and 26 $\pm$ 3.6 pM for D2L^{-/-}).

3H-SCH23390 binding experiments revealed no significant differences either in the affinity for the ligand (K_d = 197 $\pm$ 28 pM WT and 149 $\pm$ 11 pM D2L^{-/-}), nor in the number of D1R binding sites ($B_{\max}$ = 842 $\pm$ 110 WT and 779 $\pm$ 57 D2L^{-/-} fmol mg⁻¹ protein) between D2L^{-/-} and WT mice.

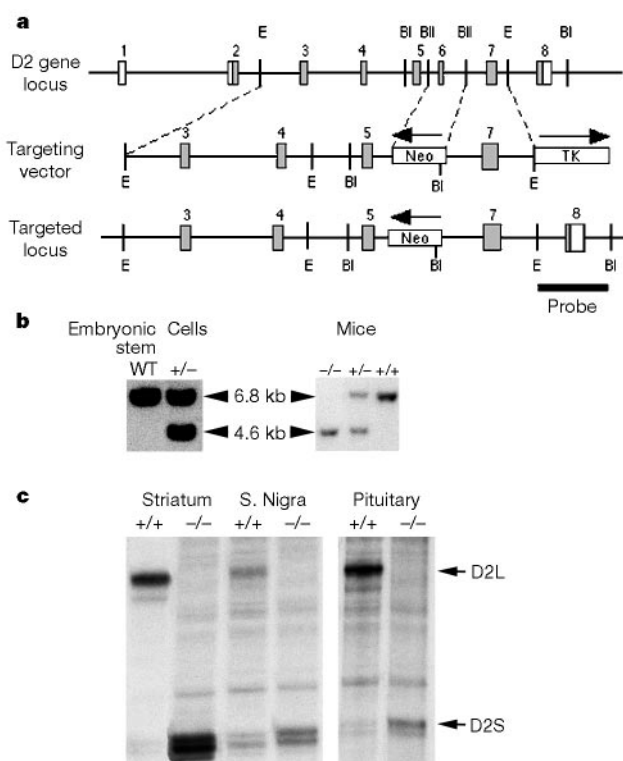


Figure 1 Disruption of D2L. **a**, D2L receptor knockout strategy. The recombinant allele was identified with a 1.4-kb *EcoRI*–*BglII* genomic probe (thick black bar). Arrows indicate the transcriptional orientation of the selection markers with respect to the D2R gene. E, *EcoRI*; BI, *BglII*; BII, *BglII*. **b**, Southern blot analysis of DNA from embryonic stem cells and tail biopsies. A 4.6-kb fragment indicates the mutant versus the WT 6.8-kb allele. **c**, RNAse protection assays of total RNA from WT and D2L^{-/-} mice. The protected fragments are 202 base pairs (bp) for D2L and 131 bp for D2S.

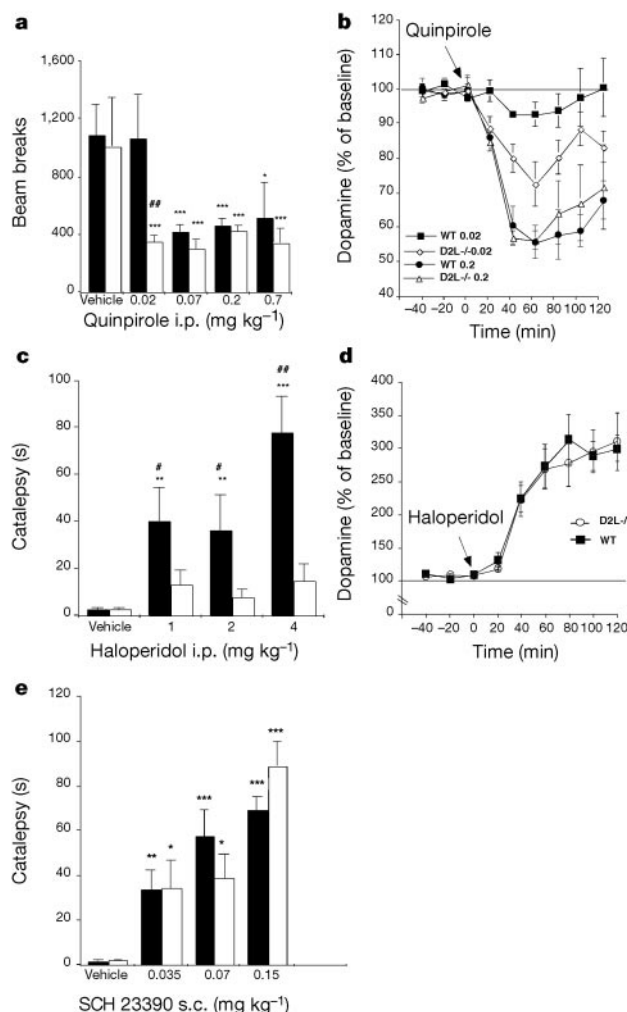


Figure 2 Responses to quinpirole, haloperidol and SCH23390 in D2L^{-/-} mice. **a**, Quinpirole effect at 0.0 mg kg⁻¹ (WT n = 13; D2L^{-/-} n = 12); at 0.02 mg kg⁻¹ (WT n = 11; D2L^{-/-} n = 10), at 0.07 mg kg⁻¹ (WT n = 5; D2L^{-/-} n = 5), at 0.2 mg kg⁻¹ (WT n = 5; D2L^{-/-} n = 6) and at 0.7 mg kg⁻¹ (WT n = 6; D2L^{-/-} n = 6). Quinpirole reduced locomotion both in WT ($F_{(4,35)} = 3.67$, $P < 0.05$) and D2L^{-/-} ($F_{(4,34)} = 3.75$, $P < 0.05$). **b**, Quinpirole (0.02 mg kg⁻¹) induced a higher decrease in striatal dopamine level in D2L^{-/-} than in WT mice (genotype effect $F_{(1,9)} = 6.01$, $P < 0.03$). 0.2 mg kg⁻¹ quinpirole similarly decreased dopamine levels in both genotypes (genotype effect $F_{(1,8)} = 0.101$, $P > 0.75$). n = 5–6 per dose per genotype. **c**, The cataleptic effects of haloperidol (n = 7 per dose per genotype) was observed only in WT (genotype effect $F_{(1,36)} = 18.42$; $P < 0.0001$). WT (dose effect $F_{(3,24)} = 5.75$, $P < 0.004$), D2L^{-/-} (dose effect $F_{(3,24)} = 1.25$, $P > 0.31$). **d**, Similarly haloperidol (0.2 mg kg⁻¹, intraperitoneally) induced increase of striatal dopamine levels in WT and D2L^{-/-} mice (n = 5 (genotype effect $F_{(1,8)} = 0.02$, $P > 0.87$). Baseline dopamine levels were: WT = 4.20 $\pm$ 0.54 pg per 40 μ l; D2L^{-/-} = 4.67 $\pm$ 0.68 pg per 40 μ l; $P > 0.58$. Equal cataleptic effect of SCH23390 in both genotypes (n = 5–8 per dose per genotype). Genotype effect $F_{(1,29)} = 0.001$; $P = 0.96$. Dose effect WT: $F_{(4,37)} = 56.381$, $P < 0.0001$; D2L^{-/-}: $F_{(4,24)} = 21.434$, $P < 0.0001$. * $P < 0.05$; ** $P < 0.01$; *** $P < 0.001$ compared to vehicle treated. * $P < 0.05$; ** $P < 0.01$ compared to the other genotype. WT, black bars; D2L^{-/-}, white bars.

D2L-null mice have intact primitive reflexes and no differences in the spontaneous locomotor behaviour were observed in comparison to WT littermates (data not shown). Similarly, no differences between genotypes were found in the locomotor activity in a novel cage and in the open field, in contrast to the phenotype of D2R^{-/-} mice².

Quinpirole, a D2-like specific agonist, was used to study the presynaptic response¹⁶ in D2L^{-/-} mice. In rodents, low doses of quinpirole induce suppression of motor activity by reducing dopamine release¹⁷. Accordingly, quinpirole induced a decrease in locomotion in both genotypes (Fig. 2a). The decrease was more pronounced in D2L^{-/-} mice, as shown by the selective effect of 0.02 mg kg⁻¹ in these mice as compared to WT (Fig. 2a).

This response strongly differs from that of D2R^{-/-} mice, in which the quinpirole effect is lost (data not shown). Thus, the preservation of D2R presynaptic function in D2L^{-/-} mice must be ascribed to D2S. Indeed, quinpirole administration results in a decrease of extracellular dopamine levels in both genotypes (Fig. 2b). However, a stronger decrease was observed in D2L^{-/-} mice at the dose of 0.02 mg kg⁻¹ with respect to WT littermates, whereas at 0.2 mg kg⁻¹ a similar decrease in dopamine was found (Fig. 2b). Thus, the neurochemical analysis perfectly mirrors the behavioural results and confirms preservation of presynaptic responses in D2L^{-/-} mice.

We then examined the effect of haloperidol, a D2-like antagonist

and antipsychotic¹. A high dose of this drug produces catalepsy¹⁸. At the neurochemical level, a low dose of haloperidol efficiently increases dopamine release¹. These two effects have been attributed to the blockade of D2 post- and presynaptic sites, respectively. We noted that in D2L^{-/-} mice, as opposed to WT littermates, haloperidol treatment did not induce a significant increase of catalepsy at any of the doses tested (Fig. 2c). A similar response is obtained in D2R-null mice¹⁸, suggesting that the D2L isoform of the D2R mediates the cataleptic effect of this antipsychotic.

Despite lack of catalepsy, haloperidol administration induced a similar increase in extracellular dopamine levels in both genotypes (Fig. 2d). These results are in line with preserved D2-dependent presynaptic function in D2L^{-/-} mice, but suggest that postsynaptic effects are impaired.

The receptor-dependency of loss of catalepsy was inferred by blockade of D1R by SCH 23390 (ref. 18), which induced catalepsy in both genotypes (Fig. 2e).

Dopaminergic postsynaptic functions in D2L^{-/-} mice were investigated by analysing the behavioural effect of three direct dopaminergic agonists: apomorphine¹⁹, a mixed D1/D2 agonist; SKF 82958 (refs 20, 21), a D1 agonist with some agonistic activity at D2 sites; and SKF 81297 (refs 20, 22), a full D1 agonist. WT animals showed a significant increase in locomotion when treated with any of these drugs (Fig. 3). In contrast, apomorphine and SKF82958 failed to induce an increase in locomotion in D2L^{-/-} mice at any dose tested (Fig. 3a, b).

The effects of SKF81297, a full D1 agonist, were also greatly diminished in D2L^{-/-} mice, although a small but statistically significant increase of locomotion was noticed at the dose of 1.5 mg kg⁻¹ (Fig. 3c). These results indicate that in D2L^{-/-} mice the locomotor-activating properties of mixed and full D1-like agonists are strongly reduced, revealing an impairment of D1 signalling. This notion is supported by the experiments performed on D2R^{-/-} mice² using this compound (Fig. 3d). SKF 81297 in these mutants elicited a significant dose-dependent increase of locomotion as in WT siblings, strongly pointing to an inhibitory control of D2S on D1-mediated activation of locomotion in D2L^{-/-} mice.

We have generated mice encoding only the D2S mRNA of D2R; this results in an invariant number of D2 sites. D2L-null mice grow and reproduce normally, which suggests a functional compensation for D2L in basal conditions. In spite of the involvement of D2R in motor functions²⁻⁴, D2L^{-/-} mice present normal spontaneous motor behaviour. In these mice, presynaptic functions related to the regulation of dopamine release and the locomotor attenuating effects of quinpirole are preserved. Similarly, extracellular dopamine levels were equally increased in WT and D2L^{-/-} mice by the administration of haloperidol. The presynaptic effects of dopaminergic ligands are probably mediated by D2S and not by other D2-like receptors, because in D2R^{-/-} mice the behavioural effects of D2/D3 selective drugs are lost²³.

Cooperative/synergistic interactions between dopamine D1 and D2 receptors have been proposed on the basis of pharmacological studies^{24,25}. The lack of D2L profoundly decreases postsynaptic responses to dopaminergic agonists, strongly suggesting that D2L is the postsynaptic D2 receptor that works in concert with D1R in producing locomotor activation. The blunted behavioural response of D2L-null mice to the D1R full agonist, SKF 81297, could result from lack of cooperation or synergy between D1/D2 pathways, owing to the absence of D2L. However, the comparison between the behavioural response of D2L^{-/-} mice, D2R^{-/-} mice and SKF 81297 (Fig. 3c and d), indicates an impairment of D1R-mediated responses. Indeed, in D2R^{-/-} mice, where both D2L and D2S are absent, SKF81297 administration produces behavioural responses similar to those of WT mice. Therefore, analysis of D2L^{-/-} mice unravels an inhibitory effect of D2S on D1R responses. The existence of postsynaptic D2 receptors that work in opposition with D1Rs has been previously proposed^{25,26}. Our results indicate

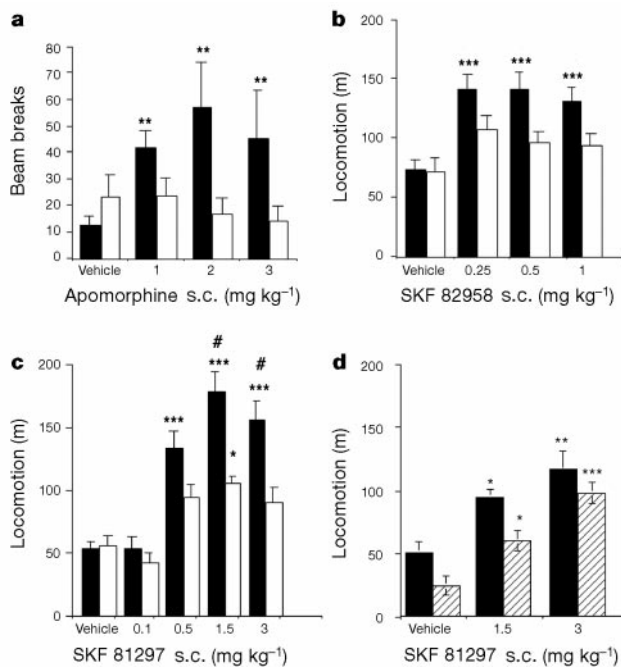


Figure 3 Locomotor response to mixed and full dopaminergic agonists in D2L^{-/-} and D2R^{-/-} mice. **a**, Apomorphine failed to induce locomotion in D2L^{-/-} mice. Genotype effect $F_{(1,24)} = 9.45$; $P < 0.006$; $n = 5$ per dose per genotype. Apomorphine effect WT (black bars): $F_{(1,12)} = 10.78$, $P < 0.007$; D2L^{-/-} (white bars): $F_{(1,12)} = 0.48$, $P > 0.48$. **b**, The locomotor response to SKF 82958 at 0.0 mg kg⁻¹ (WT $n = 18$; D2L^{-/-} $n = 17$), 0.25 mg kg⁻¹ (WT $n = 6$; D2L^{-/-} $n = 6$), 0.5 mg kg⁻¹ (WT $n = 6$; D2L^{-/-} $n = 8$) and 1 mg kg⁻¹ (WT $n = 6$; D2L^{-/-} $n = 5$). Genotype effect $F_{(1,30)} = 21$, $P < 0.0001$. SKF 82958 significantly increased locomotion in WT (black bars) (ANOVA $F_{(3,32)} = 14.476$, $P < 0.0001$), but not in D2L^{-/-} (white bars) (ANOVA $F_{(3,34)} = 2.24$, $P > 0.1$). **c**, SKF 81297 was administered at 0.0 mg kg⁻¹ (WT (black bars) $n = 19$; D2L^{-/-} (white bars) $n = 20$), 0.1, 0.5, 1.5 and 3 mg kg⁻¹ ($n = 5$ per dose per genotype). Genotype effect $F_{(1,31)} = 34.66$, $P < 0.0001$. ANOVA dose effect WT: $F_{(4,38)} = 36.63$; $P < 0.0001$; D2L^{-/-}: $F_{(4,44)} = 4.73$; $P < 0.01$. **d**, SKF 81297 effect tested in D2R^{-/-} mice (striped bars) and WT siblings (black bars). A significant increase of locomotion was found in both genotypes. Genotype effect $F_{(1,24)} = 6.87$, $P < 0.015$. ANOVA dose effect WT: $F_{(2,18)} = 9.76$, $P < 0.01$; D2R^{-/-}: $F_{(2,18)} = 20$, $P < 0.0001$. ** $P < 0.01$, *** $P < 0.001$ compared to vehicle treated; # $P < 0.05$; compared to the other genotype.

that D2S receptors might have this function *in vivo*.

Catalepsy produced by haloperidol is suppressed in D2L^{-/-} mice, a behavioural alteration very similar to what has been observed in D2R^{-/-} mice¹⁸, suggesting that D2L is the D2R targeted by haloperidol. Nevertheless, SCH 23390 (a D1R antagonist) is still able to produce catalepsy in D2L^{-/-} mice. These findings show that motor activation induced by dopamine agonists involves a strict interaction between D2 and D1 receptors, whereas motor inhibition induced by dopamine antagonists follows independent pathways.

Behavioural studies, performed on mice carrying targeted mutations, should consider the potential effect of mixed genetic backgrounds. However, we have observed preserved pharmacological D2R-mediated presynaptic responses in D2L^{-/-} mice, in spite of the absence of postsynaptic effects. Similarly, D1R-mediated functions are affected in response to agonists but not antagonists. This indicates a specific effect of the targeted mutation on the function of the dopaminergic system.

Our data identify the function of the D2L and D2S isoforms and their interactions with D1-mediated signalling; there may be potential for the modulation of dopaminergic postsynaptic responses in physiological and pathological conditions. In contrast to previous reports, D1Rs and D2Rs have recently been shown to colocalize in striatal medium spiny neurons^{27,28}. These findings imply the existence of direct intracellular cross-regulatory mechanisms between D1- and D2-activated signal transduction pathways in the striatum. Together with the differential coupling of D2L and D2S receptors to different G-proteins^{13,14}, these data suggest a possible mechanism to explain the diversity of function of these isoforms and their opposite effects on D1-activated signalling.

We propose that D2L and D2S have different and probably antagonistic functions *in vivo*. D2S is principally a D2 presynaptic²⁹ autoreceptor, which at the post-synaptic level negatively modulates D1R-dependent responses. In contrast, the D2R-mediated postsynaptic effects and their cooperative/synergistic activity with D1R seem likely to be mediated by D2L. The loss of the behavioural effects of haloperidol suggests a role for D2L in the treatment of neuropsychiatric disorders such as Tourette's syndrome and schizophrenia⁸. D2L^{-/-} mice could be used to develop drugs that can discriminate between D2S and D2L *in vivo*, leading to novel pharmacological strategies for the treatment of neuropathologies. □

Methods

Targeted mutation of D2L isoform by homologous recombination

To knock out D2L, a 5.8-kilobase (kb) *EcoRI* genomic fragment spanning exon 5 to exon 7 was subcloned into pBlueScript. From this construct a 1.5-kb *BglII* fragment containing exon 6 was deleted and replaced with the PGK-neo cassette gene. To this plasmid, a 4.8-kb *EcoRI* genomic fragment containing exon 3 and exon 4 was fused for electroporation of embryonic stem cells. The viral herpes simplex thymidine kinase (TK) gene, under the control of the PGK promoter, was included (Fig. 1a) for negative selection. The targeting vector was linearized by *NotI* before electroporation of D3 embryonic stem cells. Electroporated embryonic stem cells were selected in 150 µg ml⁻¹ G418 + 2 µM ganciclovir for 10 days. Targeted clones were identified by Southern analysis of tail genomic DNA digested with *BglII* and hybridized to a *EcoRI*-*BglII* external probe. Digestion of DNA with *BglII* yielded a 6.8-kb fragment from the WT and 4.5-kb fragment from the targeted allele.

RNAse protection and binding analyses

RNAse protections were performed as described⁶. RNAs (10 µg for brain regions and 2 µg for pituitary) were hybridized overnight at 45 °C with a molar excess of ³²P-labelled D2L-specific AccI fragment from the D2L cDNA (position 700 to 1070) linearized with *Bsu36I*. The H4 histone riboprobes was used as internal control. Ligand binding experiments were performed as described⁷.

Drugs

(-)-Quinpirole hydrochloride (LY171555); Haloperidol, SKF82958 (6-Chloro-APB HCl); SKF81297, SCH23390 and R(-)-Apomorphine hydrochloride were from Research Biochemical Incorporated.

Behavioural tests

Quinpirole (in 0.9% NaCl) was administered immediately before placing the animals in

actimetric cages and locomotion was recorded over 2 h. SKF 82958, SKF81297 (in 0.9% NaCl) and apomorphine (in 0.1% Na2S2O5) or vehicles were injected after habituation to the test cage and behaviour recorded over 1 h, using videotrack for SKFs or actimetric cages (50 × 50 cm; Imetronics) for apomorphine. Haloperidol (intraperitoneally, dissolved in a drop of glacial acetic acid and made up to volume with 0.9% NaCl) and SCH23390 (subcutaneously) were injected after habituation to the test cage (25 × 18 × 30). Catalepsy was evaluated as described¹⁸.

All values are expressed as mean ± s.e.m. Statistical analyses were performed using appropriate analysis of variance (ANOVA) followed by adapted post hoc comparisons.

Microdialysis

A guide cannula (CMA/11-Carnegie Medicin) was lowered in the striatum 2 mm above the location of the probe tip. The mouse coordinates relative to the bregma, in mm, were: A = +1, L = +1.9 and V = -2.5. Two days after recovery, microdialysis probes (CMA/11, 2-mm cuprophane membrane) were inserted. Twenty-four hours after probe implantation, animals were transferred to the microdialysis test cage (Phymep) and microdialysis (20-µl samples) performed using a full-automated online system³⁰. HPLC coupled to a coulometric detector (Coulchem II, ESA) was used to detect dopamine (detection limit: 0.3 pg per 20 µl sample). Quinpirole (0.02 and 0.2 mg kg⁻¹), and haloperidol (0.2 mg kg⁻¹) were injected intraperitoneally after three dopamine stable baseline points (less than 10% variation) were measured. Cannula placements were verified histologically.

Received 4 July; accepted 8 September 2000.

- Jackson, D. J. & Westlind-Danielsson, A. Dopamine receptors: Molecular biology, biochemistry and behavioural aspects. *Pharmacol. Ther.* **64**, 291–369 (1994).
- Baik, J. H. et al. Parkinsonian-like locomotor impairment in mice lacking dopamine D2 receptors. *Nature* **377**, 424–428 (1995).
- Kelly, M. A. et al. Locomotor activity in D2 dopamine receptor-deficient mice is determined by gene dosage, genetic background, and developmental adaptations. *J. Neurosci.* **18**, 3470–3479 (1998).
- Jung, M. et al. Potentiation of the D2 mutant motor phenotype in mice lacking dopamine D2 and D3 receptors. *Neuroscience* **91**, 911–924 (1999).
- Maldonado, R. et al. Absence of opiate rewarding effects in mice lacking dopamine D2 receptors. *Nature* **388**, 586–589 (1997).
- Saiardi, A., Bozzi, Y., Baik, J. H. & Borrelli, E. Antiproliferative role of dopamine: loss of D2 receptors causes hormonal dysfunction and pituitary hyperplasia. *Neuron* **19**, 115–126 (1997).
- Kelly, M. A. et al. Pituitary lactotroph hyperplasia and chronic hyperprolactinemia in dopamine D2 receptor-deficient mice. *Neuron* **19**, 103–113 (1997).
- Egan, M. F. & Weinberger, D. R. Neurobiology of schizophrenia. *Curr. Opin. Neurobiol.* **7**, 701–707 (1997).
- Picetti, R. et al. Dopamine D2 receptors in signal transduction and behavior. *Crit. Rev. Neurobiol.* **11**, 121–142 (1997).
- Mercuri, N. B. et al. Loss of autoreceptor function in dopaminergic neurons from dopamine D2 receptor deficient mice. *Neuroscience* **79**, 323–327 (1997).
- Calabresi, P. et al. Abnormal synaptic plasticity in the striatum of mice lacking dopamine D2 receptors. *J. Neurosci.* **17**, 4536–4544 (1997).
- L'Hirondel, M. et al. Lack of autoreceptor-mediated inhibitory control of dopamine release in striatal synaptosomes of D2 receptor-deficient mice. *Brain Res.* **792**, 253–262 (1998).
- Montmayeur, J. P., Guirmand, J. & Borrelli, E. Preferential coupling between dopamine D2 receptors and G-proteins. *Mol. Endocrinol.* **7**, 161–170 (1993).
- Guirmand, J., Montmayeur, J. P., Ceraline, J., Bhatia, M. & Borrelli, E. Alternative splicing of the dopamine D2 receptor directs specificity of coupling to G-proteins. *J. Biol. Chem.* **270**, 7354–7358 (1995).
- Crenshaw, E. B. III, Russo, A. F., Swanson, L. W. & Rosenfeld, M. G. Neuron specific alternative RNA processing in transgenic mice expressing a metallothionein-calcitonin fusion gene. *Cell* **49**, 389–398 (1987).
- Eilam, D. & Szechtman, H. Biphasic effect of D2 agonist quinpirole on locomotion and movements. *Eur. J. Pharmacol.* **161**, 151–157 (1989).
- Starke, K., Gothert, M. & Kilbinger, H. Modulation of neurotransmitter release by presynaptic autoreceptors. *Physiol. Rev.* **69**, 864–989 (1989).
- Boulay, D. et al. Haloperidol-induced catalepsy is absent in dopamine D2, but maintained in dopamine D3 receptor knock-out mice. *Eur. J. Pharmacol.* **391**, 63–73 (2000).
- Protais, P., Bonnet, J. & Constantin, J. Pharmacological characterization of the receptors involved in the apomorphine induced polyphasic modifications of locomotor activity in mice. *Psychopharmacology* **81**, 126–134 (1983).
- Ruskin, D. N., Rawji, S. S. & Walters, J. R. Effects of full D1 dopamine receptor agonists on firing rates in the globus pallidus and substantia nigra pars compacta in vivo: tests for D1 receptor selectivity and comparisons to the partial agonist SKF 38393. *J. Pharmacol. Exp. Ther.* **286**, 272–281 (1998).
- Nisenbaum, E. S., Mermelstein, P. G., Wilson, C. J. & Surmeier, D. J. Selective blockade of a slowly inactivating potassium current in striatal neurons by (±) 6-chloro-APB hydrobromide (SKF82958). *Synapse* **29**, 213–224 (1998).
- Xu, F. et al. Mice lacking the norepinephrine transporter are supersensitive to psychostimulants. *Nature Neurosci.* **3**, 465–471 (2000).
- Boulay, D., Depoortere, R., Perrault, G., Borrelli, E. & Sanger, D. J. Dopamine D2 receptor knock-out mice are insensitive to the hypolocomotor and hypothermic effects of dopamine D2/D3 receptor agonists. *Neuropharmacology* **38**, 1389–1396 (1999).
- Waddington, J. L., Daly, S. A. In *D1:D2 dopamine receptor interactions* (ed. Waddington, J.) 52–74 (Academic, London, 1993).
- Rubinstein, M., Gershnik, O. & Stefano, F. J. E. Postsynaptic bimodal effect of sulpiride on locomotor activity induced by pergolide in catecholamine-depleted mice. *Naunyn-Schmiedeberg's Arch. Pharmacol.* **337**, 115–117 (1988).
- Stoof, J. C. & Keibian, J. W. Opposing roles for D-1 and D-2 dopamine receptors in efflux of cyclic AMP from rat neo-striatum. *Nature* **294**, 366–368 (1981).
- Surmeier, D. J., Song, W. J. & Yan, Z. Coordinated expression of dopamine receptors in neostriatal medium spiny neurons. *J. Neurosci.* **16**, 6579–6591 (1996).

28. Aizman, O. *et al.* Anatomical and physiological evidence for D1 and D2 dopamine receptor colocalization in neostriatal neurons. *Nature Neurosci.* **3**, 226–230 (2000).
29. Kahn, Z. U. *et al.* Prominence of the dopamine D2 short isoform in dopaminergic pathways. *Proc. Natl Acad. Sci. USA* **95**, 7731–7736 (1998).
30. Rougé-Pont, F., Marinelli, M., LeMoal, M., Simon, H. & Piazza, P. V. Stress induced sensitization and glucocorticoids II: sensitization of the increase in extracellular dopamine induced by cocaine depends on stress-induced corticosterone secretion. *J. Neurosci.* **15**, 7189–7195 (1995).

Acknowledgements

We thank A. Saiardi for help in the initial phase of this work and D. Vallone, C. Mathis, M. Omori and J. Clifford for useful discussions. We are grateful to V. Heidt, Muriel Petit and Nelly Charrier for technical help. We thank A. Giovanni for the generous gift of SKF 81297. This work was supported by grants from INSERM, CNRS, HUS, MILDT and ARC (to E.B.), INSERM and Université de Bordeaux II (to P.V.P.) and from the Mariano Scippaccola Foundation and FRM fellowships to A.U.

Correspondence and requests for materials should be addressed to E.B. (e-mail: eb@igbmc.u-strasbg.fr).

Regulation of repulsion versus adhesion by different splice forms of an Eph receptor

Johan Holmberg, Diana L. Clarke & Jonas Frisén

Department of Cell and Molecular Biology, Medical Nobel Institute, Karolinska Institute, SE-171 77 Stockholm, Sweden

Eph tyrosine kinase receptors and their membrane-bound ephrin ligands mediate cell interactions and participate in several developmental processes^{1–4}. Ligand binding to an Eph receptor results in tyrosine phosphorylation of the kinase domain, and repulsion of axonal growth cones and migrating cells. Here we report that a subpopulation of ephrin-A5 null mice display neural tube defects resembling anencephaly in man. This is caused by the failure of the neural folds to fuse in the dorsal midline, suggesting that ephrin-A5, in addition to its involvement in cell repulsion^{5,6}, can participate in cell adhesion. During neurulation, ephrin-A5 is co-expressed with its cognate receptor EphA7 in cells at the edges of the dorsal neural folds. Three different EphA7 splice variants^{7,8}, a full-length form and two truncated versions lacking kinase domains, are expressed in the neural folds. Co-expression of an endogenously expressed truncated form of EphA7 suppresses tyrosine phosphorylation of the full-length EphA7 receptor and shifts the cellular response from repulsion to adhesion *in vitro*. We conclude that alternative usage of different splice forms of a tyrosine kinase receptor can mediate cellular adhesion or repulsion during embryonic development.

A subpopulation of ephrin-A5 null mice (17%) display severe craniofacial malformations (Fig. 1). Most affected mice lack brain and die shortly after birth. Analysis of mutant embryos revealed that this defect is caused by a failure to close the cranial neural tube. The neural plate folds correctly in ephrin-A5 null embryos (Fig. 1d). Although the neural folds become juxtaposed in the dorsal midline, they fail to fuse in some embryos and continued cell proliferation subsequently leads to eversion of the neural folds (Fig. 1f). Neural tube defects are the most common nervous system malformations in man and have a largely unknown aetiology⁹. Similar to humans, in which the incidence is higher among female fetuses¹⁰, 71% of affected ephrin-A5 null embryos were female, suggesting that mice lacking ephrin-A5 may serve as a valuable model for human neural tube defects. Folic acid supplementation reduces the risk of neural tube defects in man and certain mouse mutants^{9–11}, but failed to do so in ephrin-A5 null mice (data not shown).

To analyse the role of ephrin-A5 in neural tube closure, we first studied expression by *in situ* hybridization during neurulation (embryonic day 8–9) in wild-type embryos. Ephrin-A5 messenger RNA is expressed in the dorsal edges of the cranial neural folds before and at the time of closure (Fig. 1g, l, q). Incubation of embryos with ephrin-A5-Fc, to visualize receptor expression, failed to detect binding at the 10-somite stage (Fig. 1h), but it was strong at the time of fusion (Fig. 1m). Polymerase chain reaction with reverse transcription (RT–PCR) analysis of the edges of the cranial neural folds with specific primers for all known EphA receptors revealed expression of only EphA7 mRNA (data not shown). The EphA7 locus encodes one full-length (FL) receptor and two truncated receptors (T1 and T2) lacking the kinase domain^{7,8}. RT–PCR, *in situ* hybridization and immunohistochemistry showed expression of all three splice variants in the edges of the cranial neural folds (Fig. 1), which precedes, but otherwise mimics, the pattern of ephrin-A5-Fc labelling. That EphA7 is the main functional ephrin-A5 receptor during neurulation is supported by a similar penetrance (24%) of anencephaly in EphA7 null mice (J.H., T. Ciossek, A. Ullrich and J.F., unpublished data). EphA7-Fc binding was prominent at the edges of the neural folds in wild-type embryos (Fig. 1t), but corresponding labelling was absent in ephrin-A5 null embryos (data not shown), indicating that ephrin-A5 is the only EphA7 ligand expressed in the cranial neural folds. Both ephrin-A5 and EphA7 expression is restricted to the cranial level of the neural folds, corresponding to the strict localization of the neural tube defects to this level.

It seems surprising that cells that adhere to each other during neurulation express these molecules with well-characterized repulsive properties, and that ephrin-A5 and EphA7 are apparently important for adhesion in this context, as neural tube fusion is impaired in their absence. There are a few additional situations in which the role of ephrins cannot easily be explained by a repulsive action. For example, ephrin-A1 has chemoattractant effects on endothelial cells¹². Furthermore, whereas many axons are repelled by ephrin-A5 *in vitro*, ephrin-A5 induces growth of axons from certain sympathetic and cortical neurons^{13,14}, and may mediate attachment of fibroblasts¹⁵. Moreover, the cleft palate observed in EphB2/EphB3 double mutant mice¹⁶ is difficult to explain by loss of repulsive ephrin activity. To test directly the role of ephrin-A5 in cell adhesion during neurulation, we analysed the adhesion of cells dissected from the edges of the cranial neural folds from wild-type, ephrin-A5^{+/-} and ephrin-A5^{-/-} embryos to EphA7-Fc coated wells. We found an inverse correlation between ephrin-A5 gene dosage and adhesion (Fig. 2a), and a statistically significant decrease in adhesion of ephrin-A5 mutant cells as compared with wild-type cells ($P < 0.01$, $n = 3–9$ embryos).

Tyrosine kinase receptors dimerize or oligomerize on ligand binding, enabling cross-phosphorylation and signal transduction¹⁷, and mutant tyrosine kinase receptors lacking kinase activity can act as dominant-negative inhibitors of signalling¹⁷. Moreover, injection of RNA encoding experimentally truncated versions of Eph receptors into *Xenopus* and zebrafish embryos disrupts patterning^{18,19}, although the mechanism for this has not yet been analysed. We therefore determined whether expression of an endogenously expressed truncated EphA7 receptor could alter the response of EphA7-FL-expressing cells to ephrin-A5. To study ephrin-A5 and EphA7 interactions in 293 cells, which lack endogenous type A ephrins and Eph receptors²⁰, we generated stable 293 lines constitutively expressing EphA7-FL, and with a doxycyclin-inducible expression of EphA7-T1 (Fig. 2b). We first analysed whether expression of the truncated receptor would alter the repulsive effect of ephrin-A5 binding to EphA7-FL in a transwell chemotaxis assay. Whereas 293 cells expressing EphA7-FL were significantly repelled by clustered ephrin-A5-Fc as compared with wild-type 293 cells ($P < 0.004$), there was no statistically significant difference in repulsion between cells co-expressing EphA7-FL and EphA7-T1 or

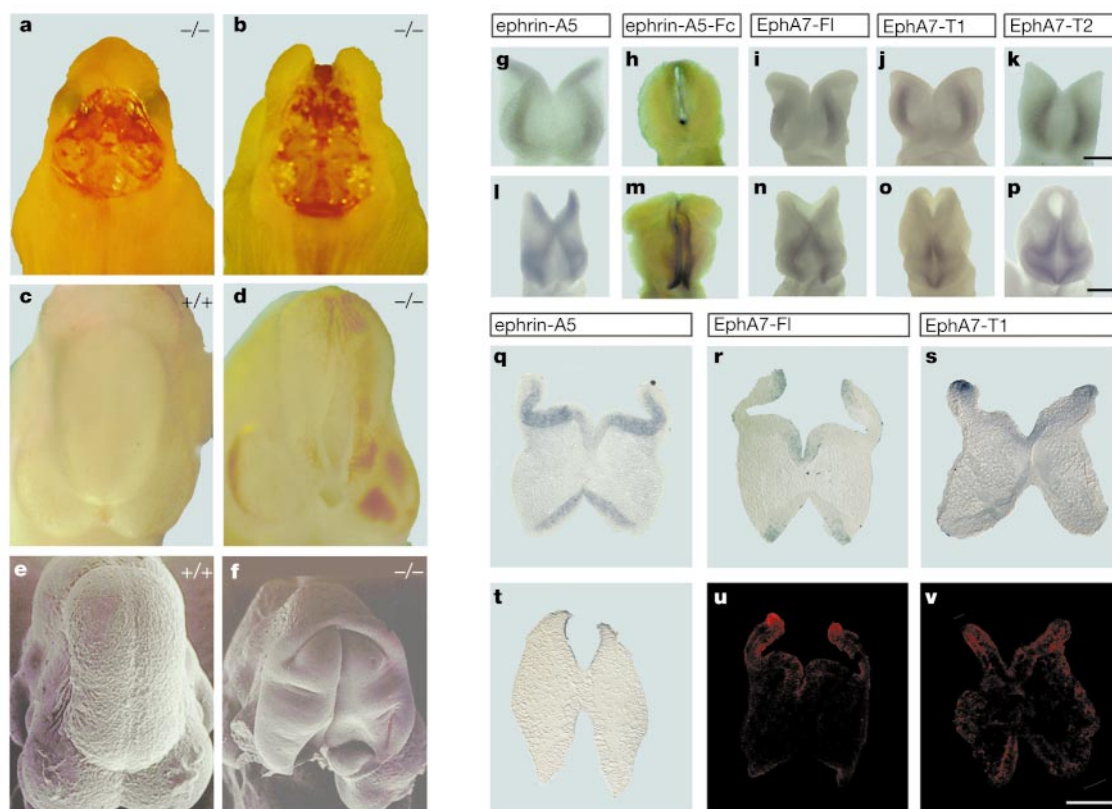


Figure 1 Neural tube defects in ephrin-A5 null mice. A subset of ephrin-A5^{-/-} mice display anencephaly (**a**) sometimes with cleft nose and palate (**b**). This is due to the cranial neural folds failing to fuse in the dorsal midline (**c–f**). Embryos are 10 days in **c** and **d**; 10.5 days in **e** and **f**. *In situ* hybridization localization of ephrin-A5 (**g**, **i**, **q**), EphA7-FL (**i**, **n**, **r**), EphA7-T1 (**j**, **o**, **s**), EphA7-T2 (**k**, **p**) mRNA, ephrin-A5-Fc (**h**, **m**) and EphA7-Fc (**t**)

binding; and immunofluorescence localization of EphA7-FL (**u**) and EphA7-T1 (**v**) in the head region at somite stages 9–11 (**g–k**), 11–14 (**q–v**) and 14–16 (**l–p**). **g–p** show whole-mount labelling and **q–v** horizontal cross-sections through the head region (ventral is up). Scale bars, 200 μ m.

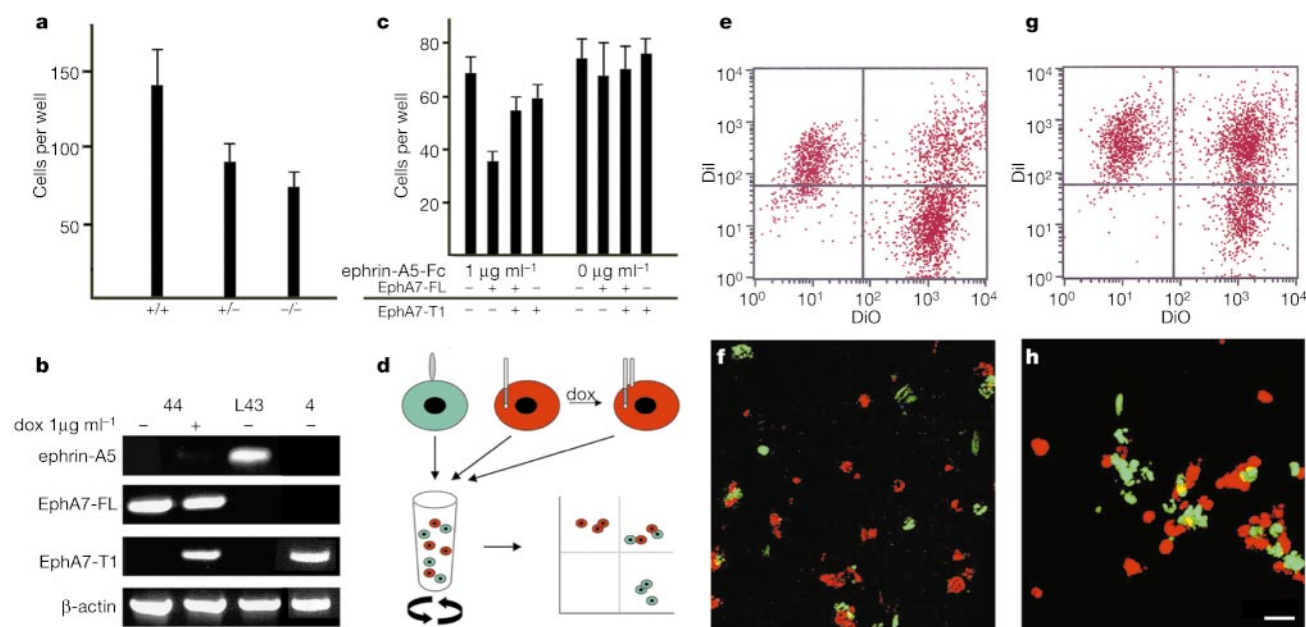


Figure 2 Ephrin-A5 and EphA7 splice versions in cell adhesion and migration. **a**, Adhesion of neural fold cells from day 8.5 embryos with different ephrin-A5 gene dosage to EphA7-Fc coated wells (mean $\pm$ s.e.m.). **b**, RT-PCR analysis of stable cell lines (**b**), which first were used to study the influence of EphA7 splice form expression on cell repulsion in response to clustered ephrin-A5-Fc in a chemotaxis chamber (**c**). The number of receptor-expressing cells that had migrated towards the ligand is shown in (**c**)

(mean $\pm$ s.e.m., $n = 3–5$). **d**, The effect of expression of different EphA7 splice forms on adhesion to ephrin-A5-expressing cells was next tested in an aggregation assay. The cells were labelled with Dil (red, receptor-expressing cells) or DiO (green, ligand-expressing cells). **e–h**, FACS analysis and photomicrographs of uninduced (**e**, **f**) and doxycyclin-induced (**g**, **h**) receptor-expressing cells, respectively, after aggregation with ephrin-A5-expressing cells. Scale bar in **h**, 20 μ m.

expressing EphA7-T1 only as compared with wild-type cells (Fig. 2c). We next determined whether EphA7-T1 could alter the adhesive response of EphA7-FL-expressing cells to ligand binding, and carried out aggregation assays with 293 cells stably expressing ephrin-A5 (Fig. 2d–h). The number of mixed cell aggregates containing both ligand- and receptor-expressing cells was quantified by fluorescence-activated cell sorting (FACS) analysis. Induction of EphA7-T1 expression significantly ($P < 0.00005$) increased the number of mixed aggregates (Fig. 2g, h); that is, it augmented adhesion between cells expressing ligand and cells expressing receptor. There were 1.6 ± 0.05 times more cell aggregates containing both receptor- and ligand-expressing cells when EphA7-T1 expression was induced in EphA7-FL cells (mean $\pm$ s.e.m. from seven independent experiments).

As the aggregation assay analyses cell–cell interactions in solution, the repulsion might be ineffective because of the lack of a stable substrate for the cells. To assay cellular repulsion, therefore, we co-cultured ligand- and receptor-expressing cells. When cells expressing only EphA7-FL were cultured with ephrin-A5-expressing cells, they clearly avoided the ligand-expressing cells (Fig. 3a). However, induction of EphA7-T1 expression in the EphA7-FL cells made these cells mix with, rather than avoid, the ligand-expressing cells (Fig. 3b). This suggests that the presence of EphA7-T1 shifted the cellular response of the ephrin/Eph receptor interaction from repulsion to adhesion. However, it is possible that EphA7-T1 merely suppressed repulsion and that the mixing with ligand-expressing cells is independent of ephrin-A5 interaction with EphA7. To test whether expression of EphA7-T1 together with EphA7-FL resulted in adhesion to ligand-expressing cells, we tested the preference of the receptor-expressing cells to grow on

ephrin-A5-expressing or wild-type 293 cells. We plated the receptor-expressing cells on preformed confluent mosaics of ephrin-A5 and wild-type cells. We found that the EphA7-FL-expressing cells preferred to grow on the wild-type cells (Fig. 3c, e, f), whereas the cells co-expressing both receptor forms grew predominantly on the ephrin-A5-expressing cells ($P < 0.000005$), as did cells expressing only EphA7-T1 ($P < 0.003$) (Fig. 3d–f). These results show that expression of the T1 truncated splice form of EphA7 turns the repulsive response to ephrin-A5 into adhesion.

The repulsive effect of Eph receptor signalling is dependent on phosphorylation of the kinase domain, which appears to induce cytoskeletal rearrangements and decrease integrin-mediated adhesion^{21,22}. To determine whether EphA7-T1 can block the phosphorylation of the EphA7-FL receptors, we assayed tyrosine phosphorylation of EphA7-FL in response to ligand binding. Co-expression of EphA7-T1 substantially decreased tyrosine phosphorylation of EphA7-FL (Fig. 4a). On the basis of this, we propose a model where co-expression of EphA7-T1 inhibits repulsive signal transduction by EphA7-FL and enables an adhesive interaction (Fig. 4c). In the closing neural tube, both receptor forms as well as the ligand are expressed (Figs. 1, 4b), suggesting that this model (Fig. 4c) might explain the loss of adhesion between the neural folds in the ephrin-A5 null mice.

Because both ephrins and Eph receptors are membrane anchored, suppression of repulsive signal transduction could hypothetically turn these proteins into cell-adhesion molecules. Alternatively, the adhesive effect may be indirect where the ligand–receptor interaction may result in kinase-independent signal transduction that could affect other molecules with adhesive properties.

Notably, mutations in the gene encoding the *Caenorhabditis elegans* Eph receptor, VAB-1, result in defective ventral enclosure, a process resembling neurulation²³. Analysis of different *vab-1* mutants revealed that this process is independent of VAB-1 kinase activity²³, suggesting that the kinase-independent adhesive functions of Eph family receptors may be evolutionarily conserved.

Over the past few years it has become clear that several molecules involved in axon guidance and cell migration, including netrins, semaphorins and slits, have both attractive and repulsive actions. The dual function of these molecules is regulated by interaction with different receptors or by modulation of cyclic-nucleotide-dependent signal transduction pathways^{24,25}. The differential response to ephrin-A5 shows, to our knowledge, the first example of how the response to a ligand can be modulated between repulsion

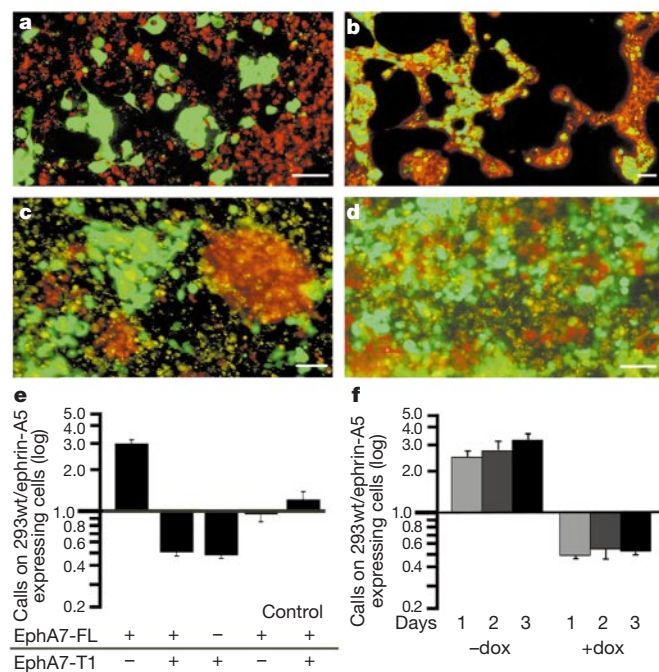


Figure 3 Regulation of repulsion versus adhesion by different EphA7 splice forms. **a, b**, EphA7-FL-expressing cells (red) avoid ephrin-A5-expressing cells (green) (**a**), but grow closely together with ephrin-A5-expressing cells when EphA7-T1 is induced (**b**). **c, d**, Uninduced (**c**) or doxycycline-induced (**d**) receptor-expressing cells are plated on a preformed confluent mosaic of wild-type 293 cells (DiA labelled, yellow) and ephrin-A5-expressing 293 cells (green fluorescent protein expressing). **e**, Ratio of receptor-expressing cells growing on wild-type (wt) 293/ephrin-A5-expressing cells (mean $\pm$ s.e.m., $n = 4-5$). In control experiments, receptor-expressing cells are cultured on confluent mosaics of dye labelled wild-type cells ($n = 3$). **f**, Distribution of receptor-expressing cells over time after plating on the mosaic ($n = 3$). Scale bars, 30 μ m.

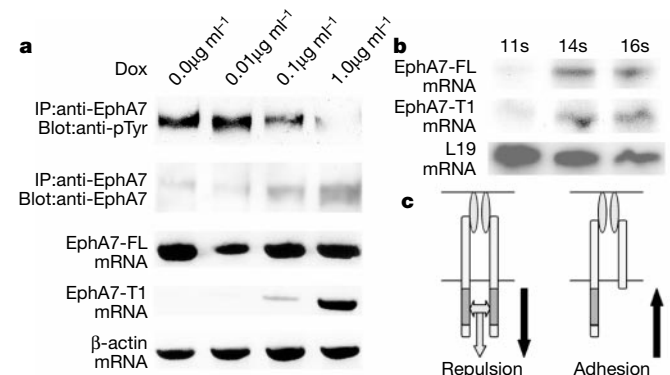


Figure 4 EphA7-T1 suppresses tyrosine phosphorylation of EphA7-FL. **a**, Dose-dependent suppression by EphA7-T1 of EphA7-FL tyrosine phosphorylation in response to contact with ephrin-A5-expressing cells. Levels of EphA7-FL and EphA7-T1 mRNA are shown by RT-PCR. **b**, Levels of EphA7-FL and EphA7-T1 mRNA in the neural folds at somite stages 11, 14 and 16 analysed by RT-PCR. **c**, Model for how EphA7-T1 might suppress EphA7-FL phosphorylation and inhibit the signal transduction that results in cell repulsion, and so instead allow an adhesive interaction.

or adhesion by alternative use of different splice forms of a receptor. □

Methods

Analysis of mutant mice

We genotyped ephrin-A5 mutant mice by PCR as described⁶. For scanning electron microscopy, embryos were fixed and, following dehydration, dried in hexamethylsilazane, mounted with silver paste on aluminium stubs, gold/palladium coated and examined in a Zeiss DSM 950 scanning electron microscope.

In situ hybridization and protein localization

Digoxigenin-labelled riboprobes complementary to *ephrin-A5* (ref. 6), *EphA7-FL* (bases 2,601–2,925), *EphA7-T1* (2,017–2,486) and *EphA7-T2* (2,019–2,294)⁷ were used for *in situ* hybridization as described²⁶. We carried out ephrin-A5-Fc and EphA7-Fc binding as described²⁷. EphA7-FL and EphA7-T1 were localized by immunofluorescence using splice-form-specific antibodies^{7,28}.

RT-PCR

Total RNA was isolated from the neural folds of embryos or cell lines with the RNeasy kit (Qiagen). Reverse transcription was performed with superscript-II and the cDNA was amplified by PCR with primers specific for ephrin-A5, 5'-ATT TGC GGC CGC TTC ATC GTA TTT ATT CCT ATT TAT T-3' and 5'-GCT CGA GGG GTT GCT GCT GTT CCA GTA GAC GG-3'; EphA7-FL, 5'-ATT TGC GGC CGC TCT ACA CCA CGA CTG GTG GAA AAA T-3' and 5'-CCG CTC GAG CTT GGG TTT CGA ATC ATT TTG TCT A-3'; EphA7-T1, 5'-ACT CTA CTT TCA TTC TTT AGT AAA-3' and 5'-GAA GCA TGC AGG TAA CAC ATA AAT G-3'; EphA7-T2, 5'-ATT TGC GGC CGC ACT TCT ACT TTC ATT CTC TTT ACA GGG A-3' and 5'-CCG CTC GAG CTG AGC ACT AAT TAT CAA AAA AGA T-3'.

Cell lines

We transfected *EphA7-FL* complementary DNA in pZeoSV2 under Zeosin selection into 293 cells stably expressing rTIA. We transfected these cells with pTK-Hyg and *EphA7-T1* cDNA in pTRE, selected with hygromycin, and identified several cell lines with doxycycline-inducible ($1 \mu\text{g ml}^{-1}$ for 48 h) expression of EphA7-T1 and stable expression of EphA7-FL. We generated additional stably transfected clonal 293 cell lines with an *ephrin-A5* cDNA in pREX2N or *EphA7-T1* cDNA in pZeoSV2. The different cell lines showed indistinguishable responses in the assays.

Adhesion, chemotaxis, aggregation and co-cultivation assays

We dissected the edges of the cranial neural folds from 8.5-day-old embryos from ephrin-A5^{+/+} intercrosses. Equal numbers of cells from individual embryos were dissociated in Ca²⁺- and Mg²⁺-free PBS, and cells from single embryos were plated in EphA7-Fc-coated ($5 \mu\text{g ml}^{-1}$) wells on a rotary shaker (100 r.p.m. at 37 °C) and allowed to adhere for 1 h. The wells were rinsed to remove non-adherent cells; the remaining cells were fixed and the number of EphA7-FL-immunoreactive cells (used here as a marker for cells from the edge of the neural folds) in each well was quantified. The adhesion assays were carried out blindly of genotype, which was established by PCR of DNA from the individual embryos.

The chemotaxis assay was performed essentially as described²⁹. Briefly, 27 μl of medium or medium containing EphA7-Fc fusion protein ($5 \mu\text{g ml}^{-1}$) clustered with anti-human immunoglobulin- γ (IgG) antibodies ($50 \mu\text{g ml}^{-1}$) was added to the lower wells. The upper wells were then filled with 2×10^5 293 cells per ml.

For the aggregation assay, cells were labelled by addition of DiI (0.33 ng ml^{-1}) or DiO (1.7 ng ml^{-1}) to the tissue culture medium 48 h before the assay or by stable transfection with pEGFP-1. Cells (10^7) were transferred to polystyrene tubes in 1 ml of buffer (137 mM NaCl, 4.7 mM KCl, 0.6 mM MgSO₄, 1.2 mM CaCl₂, 10 mM Hepes, pH 7.4) with 1 mg ml^{-1} glucose and 20 U ml⁻¹ Dnase³⁰. The tubes were placed on a rotary shaker (200 r.p.m. for 1 h at 37 °C) and aggregate formation was analysed in a FACScalibur flow cytometer (Becton Dickinson) equipped with Cellquest 3.1 software.

For the co-cultivation experiments, we mixed and plated labelled cells ($\text{DiI } 1 \text{ ng ml}^{-1}$, $\text{DiA } 5 \text{ ng ml}^{-1}$, $\text{DiO } 5 \text{ ng ml}^{-1}$ or stably pEGFP-1 transfected) together in tissue culture plates and allowed them to grow together for 12–24 h. In other co-cultivation experiments, wild-type 293 cells (labelled with DiA) and pEGFP-1 transfected cells expressing ephrin-A5 were plated together and allowed to grow to confluency, creating a mosaic monolayer. Induced and uninduced receptor-expressing cells were labelled with DiI, plated on the monolayer and their distribution was analysed 12–24 h later by fluorescence microscopy. The ratio of receptor-expressing cells growing on wild-type or ligand-expressing cells was analysed by image analysis using Adobe Photoshop 4.0. We carried out all analyses of statistical significance by Student's *t*-test.

Phosphorylation assay

Cells were grown to confluency in 100-mm tissue culture dishes. Ephrin-A5-expressing cells, or wild-type 293 cells in control experiments, were plated on the confluent cell layer to stimulate phosphorylation of the Eph receptors, and were washed away after 10 min. Cells were lysed in RIPA buffer and the protein content between samples was equalized. Lysates were incubated overnight at 4 °C with rabbit anti-EphA7 antiserum (1/1,000)⁷ and were then adsorbed for 2 h at 4 °C with 100 μl of 50% Sepharose-G. After washing, the samples were boiled, resolved on 8% SDS polyacrylamide gels, transferred to nitrocellulose filters which were incubated with mouse monoclonal anti-phosphotyrosine (4G10, $1 \mu\text{g ml}^{-1}$; Upstate Biotechnology).

Received 9 June; accepted 31 July 2000.

- Flanagan, J. G. & Vanderhaeghe, P. The ephrins and Eph receptors in neural development. *Annu. Rev. Neurosci.* **21**, 309–345 (1998).
- Frisén, J., Holmberg, J. & Barbacid, M. Ephrins and their eph receptors: multitasking directors of embryonic development. *EMBO J.* **18**, 5159–5165 (1999).
- Holder, N. & Klein, R. Eph receptors and ephrins: effectors of morphogenesis. *Development* **126**, 2033–2044 (1999).
- O'Leary, D. D. M. & Wilkinson, D. G. Eph receptors and ephrins in neural development. *Curr. Opin. Neurobiol.* **9**, 65–73 (1999).
- Drescher, U. *et al.* In vitro guidance of retinal ganglion cell axons by RAGS, a 25 kDa tectal protein related to ligands for Eph receptor tyrosine kinases. *Cell* **82**, 359–370 (1995).
- Frisén, J. *et al.* Ephrin-A5 (AL-1/RAGS) is essential for proper retinal axon guidance and topographic mapping in the mammalian visual system. *Neuron* **20**, 235–243 (1998).
- Ciossek, T., Millauer, B. & Ullrich, A. Identification of alternatively spliced mRNAs encoding variants of MDK1, a novel receptor tyrosine kinase expressed in the murine nervous system. *Oncogene*, **10**, 97–108 (1995).
- Valenzuela, D. M. *et al.* Identification of full-length and truncated forms of Ehk-3, a novel member of the Eph receptor tyrosine kinase family. *Oncogene* **10**, 1573–1580 (1995).
- Botto, L. D., Moore, C. A., Khoury, M. J. & Erickson, J. D. Neural-tube defects. *N. Engl. J. Med.* **341**, 1509–1519 (1999).
- Källén, B. *et al.* International study of sex ratio and twinning of neural tube defects. *Teratology* **50**, 322–331 (1994).
- Fleming, A. & Copp, A. J. Embryonic folate metabolism and mouse neural tube defects. *Science* **280**, 2107–2109 (1998).
- Pandey, A., Shao, H., Marks, R. M., Polverini, P. J. & Dixit, V. M. Role of B61, the ligand for the Eck receptor tyrosine kinase, in TNF- α -induced angiogenesis. *Science* **268**, 567–569 (1995).
- Castellani, V., Yue, Y., Gao, P.-P., Zhou, R. & Bolz, J. Dual action of a ligand for Eph receptor tyrosine kinases on specific populations of axons during the development of cortical circuits. *J. Neurosci.* **18**, 4663–4672 (1998).
- Gao, P.-P., Sun, C.-H., Zhou, X.-F., DiCicco-Bloom, E. & Zhou, R. Ephrins stimulate or inhibit neurite outgrowth and survival as a function of neuronal cell type. *J. Neurosci. Res.* **60**, 427–436 (2000).
- Davy, A. *et al.* Compartmentalized signaling by GPI-anchored ephrin-A5 requires the Fyn tyrosine kinase to regulate cellular adhesion. *Genes Dev.* **13**, 3125–3135 (1999).
- Orioli, D., Henkemeyer, M., Lemke, G., Klein, R. & Pawson, T. Sek4 and Nuk receptors cooperate in guidance of commissural axons and in palate formation. *EMBO J.* **15**, 6035–6049 (1996).
- van der Geer, P., Hunter, T. & Lindberg, R. A. Receptor protein-tyrosine kinases and their signal transduction pathways. *Annu. Rev. Cell Biol.* **10**, 251–337 (1994).
- Xu, Q. L., Alludus, G., Holder, N. & Wilkinson, D. G. Expression of truncated Sek-1 receptor tyrosine kinase disrupts the segmental restriction of gene expression in the *Xenopus* and zebrafish forebrain. *Development* **121**, 4005–4016 (1995).
- Xu, Q., Alludus, G., Macdonald, R., Wilkinson, D. G. & Holder, N. Function of the Eph-related kinase *rtk1* in patterning of the zebrafish forebrain. *Nature* **381**, 319–322 (1996).
- Nakamoto, M. *et al.* Topographically specific effects of Elf-1 on retinal axon guidance in vitro and retinal mapping in vivo. *Cell* **86**, 755–766 (1996).
- Holland, S. J., Peles, E., Pawson, T. & Schlessinger, J. Cell-contact-dependent signalling in axon growth and guidance: Eph receptor tyrosine kinases and receptor protein tyrosine phosphatase beta. *Curr. Opin. Neurobiol.* **8**, 117–127 (1998).
- Zou, J. X. *et al.* An Eph receptor regulates integrin activity through R-Ras. *Proc. Natl. Acad. Sci. USA* **96**, 13813–13818 (1999).
- George, S. E., Simokat, K., Hardin, J. & Chisholm, A. D. The VAB-1 Eph receptor tyrosine kinase functions in neural and epithelial morphogenesis in *C. elegans*. *Cell* **92**, 633–643 (1998).
- Song, H. *et al.* Conversion of neuronal growth cone response from repulsion to attraction by cyclic nucleotides. *Science* **281**, 1515–1518 (1998).
- Hong, K. *et al.* A ligand-gated association between cytoplasmic domains of UNC5 and DCC family receptors converts netrin-induced growth cone attraction to repulsion. *Cell* **97**, 927–941 (1999).
- Henrique, D. Expression of a *Delta* homologue in prospective neurons in the chick. *Nature* **375**, 787–790 (1995).
- Cheng, H.-J. & Flanagan, J. G. Identification and cloning of Elf-1, a developmentally expressed ligand for the Mek4 and Sek receptor tyrosine kinases. *Cell* **79**, 157–168 (1994).
- Ciossek, T., Ullrich, A., West, E. & Rogers, J. H. Segregation of the receptor EphA7 from its tyrosine kinase-negative isoform on neurons in adult mouse brain. *Brain Res. Mol. Brain Res.* **74**, 231–236 (1999).
- Wang, H. U. & Anderson, D. J. Eph family transmembrane ligands can mediate repulsive guidance of trunk neural crest migration and motor axon outgrowth. *Neuron* **18**, 383–396 (1997).
- Wikström, K., Kjellström, G. & Öbrink, B. Homophilic intercellular adhesion mediated by C-CAM is due to a domain 1-domain 1 reciprocal binding. *Exp. Cell Res.* **227**, 360–366 (1996).

Acknowledgements

We thank M. Barbacid, J. Ericson, J. Flanagan, U. Lendahl, B. Öbrink, L. Philipson, D. Wilkinson and members of our laboratory for helpful discussions; S. Park for help with vector construction; T. Ciossek for EphA7 cDNAs and antisera; J. Wasserman for assistance with scanning electron microscopy; and B. Singer and G. Greicius for help with FACS analysis. This study was supported by grants from the Swedish Foundation for Strategic Research, the Karolinska Institute, the Swedish Medical Research Council, the Swedish Cancer Society, the Swedish Medical Society, Hedlunds stiftelse, Ostermans stiftelse, Magnus Bergvalls stiftelse, Marcus Borgströms stiftelse, Hagbergs stiftelse, Tore Nilssons stiftelse, the Wenner-Gren foundation, Åke Wibergs stiftelse and Kapten Arthur Erikssons fond. D.L.C. was supported by a Wenner-Gren postdoctoral fellowship.

Correspondence and requests for materials should be addressed to J.F. (e-mail: jonas.frisen@cmb.ki.se).

Functional interaction of phytochrome B and cryptochrome 2

Paloma Más, Paul F. Devlin, Satchidananda Panda & Steve A. Kay

Department of Cell Biology and National Science Foundation Center for Biological Timing, The Scripps Research Institute, 10550 North Torrey Pines Road, La Jolla, California 92037, USA

Light is a crucial environmental signal that controls many photomorphogenic and circadian responses in plants¹. Perception and transduction of light is achieved by at least two principal groups of photoreceptors, phytochromes and cryptochromes^{2,3}. Phytochromes are red/far-red light-absorbing receptors encoded by a gene family of five members (phyA to phyE)^{2,4} in *Arabidopsis*. Cryptochrome 1 (cry1), cryptochrome 2 (cry2) and phototropin are the blue/ultraviolet-A light receptors that have been characterized in *Arabidopsis*⁵. Previous studies showed that modulation of many physiological responses in plants is achieved by genetic interactions between different photoreceptors⁶; however, little is known about the nature of these interactions and their roles in the signal transduction pathway. Here we show the genetic inter-

action that occurs between the *Arabidopsis* photoreceptors phyB and cry2 in the control of flowering time, hypocotyl elongation and circadian period by the clock. PhyB interacts directly with cry2 as observed in co-immunoprecipitation experiments with transgenic *Arabidopsis* plants overexpressing cry2. Using fluorescent resonance energy transfer microscopy, we show that phyB and cry2 interact in nuclear speckles that are formed in a light-dependent fashion.

Light signals markedly affect the pace of the clock by activating different photoreceptors⁷. In response to increasing light intensity, the period length of the circadian expressed *CAB2::luciferase* gene⁸ decreases⁷. Here, *cry2* mutant seedlings displayed a deficiency in the perception of white light, as shown by a longer period length of *CAB2::luciferase* expression at intermediate fluence rates (Fig. 1a). This effect was not observed in red light (Fig. 1b), whereas in blue light only a small increment was observed at high fluence rates (Fig. 1c). These results indicate that full *cry2* function is only apparent on simultaneous illumination with multiple wavelengths of light, and suggests that it has a requirement for phytochrome activation. The significance of the shorter period length in *cry2* mutants at the lowest red or blue fluence rates remains unclear.

Light with a reduced red/far-red (R/FR) ratio that mimics light reflected from neighbouring vegetation causes the conversion of phytochrome from the active Pfr form to the inactive Pr form, and results in a shade-avoidance response including promotion of elongation growth and accelerated flowering⁹. *Cry2*-deficient

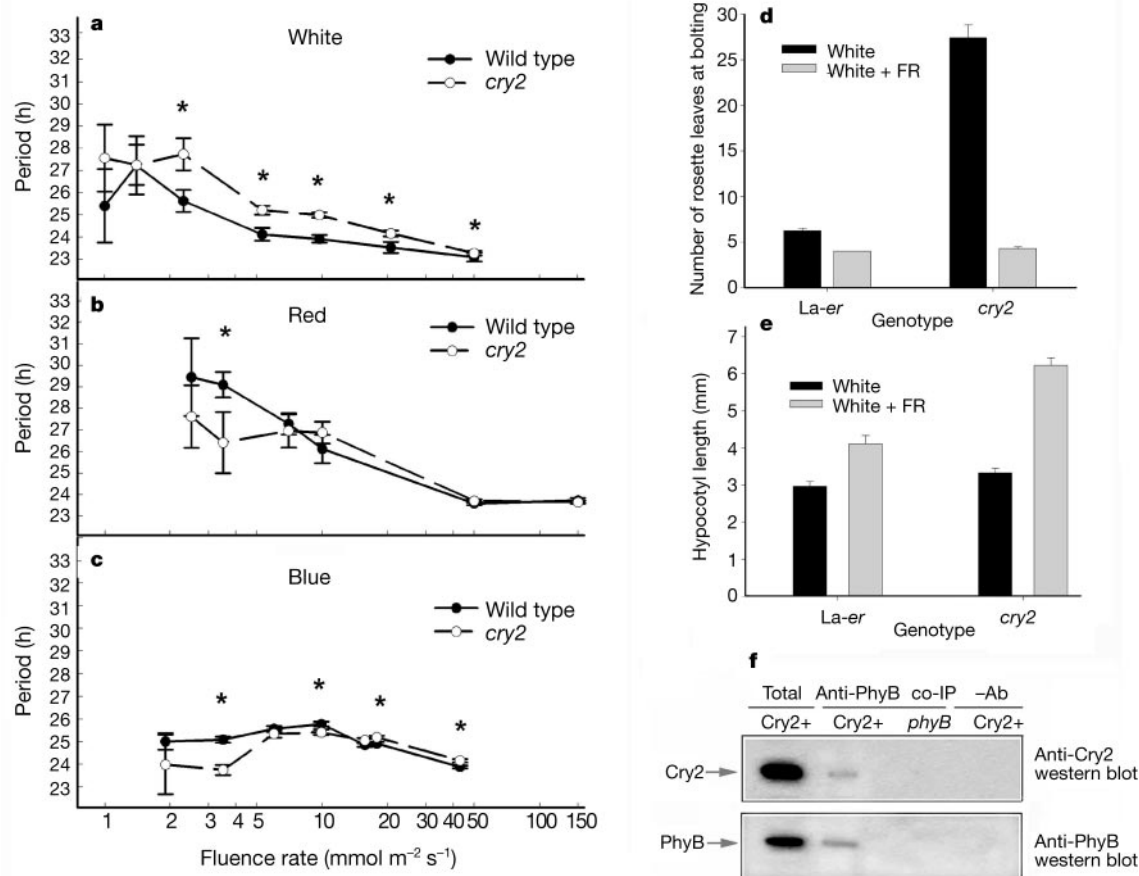


Figure 1 Physiological and biochemical evidence for interaction between *cry2* and phyB. **a–c**, Effect of light intensity on mean period length ($\pm$ s.e.) of the circadian rhythm of *CAB2::LUC* bioluminescence in wild type and *cry2* seedlings in white light (**a**), red light (**b**) or blue light (**c**). Asterisk, $P < 0.01$ (Student's two tail heteroscedastic *t*-test). **d**, Mean flowering time ($\pm$ s.e.) in wild-type and *cry2* seedlings in high (white) versus low (white + FR) R/FR ratio. **e**, Mean hypocotyl length ($\pm$ s.e.) in wild-type and *cry2* seedlings in

high (white) versus low (white + FR) R/FR ratio. **f**, Western blot analysis of co-immunoprecipitation (co-IP) experiments with anti-phyB antibody and detection with antibody to cry2. The membrane was stripped and reprobed with anti-phyB antibody. 'Total' indicates total protein extracts not subjected to immunoprecipitation. *cry2+*, *cry2* overexpression *Arabidopsis* lines; *phyB*, *phyB* mutant *Arabidopsis* lines. –Ab indicates samples processed without incubation with phyB antibody.

Arabidopsis plants have a delayed flowering phenotype in response to extended photoperiods¹⁰, and regulation of floral induction is mediated by the antagonistic actions of *cry2* and *phyB*¹¹. Here, the late flowering phenotype of *cry2* in white light was no longer apparent under a low R/FR ratio (white + FR light), showing a marked acceleration of flowering time (Fig. 1d). These results indicate that the active Pfr form of phytochrome is required for the expression of *cry2* phenotype. Thus, either *cry2* suppresses *phyB* signalling, or removal of Pfr overrides other pathways in which *cry2* regulates flowering. Furthermore, although *cry2* seedlings showed a wild-type hypocotyl length in white light, they displayed a long hypocotyl phenotype in white + FR light (Fig. 1e). We propose that long hypocotyl phenotype of *cry2* is not apparent in white light because it is negated by increased *phyB* action in the absence of *cry2* (*phyB* is the major phytochrome that inhibits hypocotyl elongation in white light¹²). In agreement with this, *cry2* seedlings showed a

greater promotion of elongation growth than wild-type seedlings. These results are consistent with a suppression of *phyB* signalling by *cry2*. Alternatively, a conditional redundancy between *cry2* and phytochrome signalling may also be possible.

To examine whether *phyB* directly interacts with *cry2*, we performed co-immunoprecipitation assays with *Arabidopsis* transgenic plants overexpressing *cry2* (*cry2+*)¹³. Immunoprecipitation with anti-*phyB* antibody and subsequent detection of *cry2* (Fig. 1f, anti-*cry2* western blot) showed a band with a relative molecular mass of about 75,000 ($M_r \approx 75K$) (lane co-IP, *cry2+*) coincident with the size of the band observed in total protein extracts not subjected to immunoprecipitation (lane total, *cry2+*). The absence of signal in *phyB*-deficient *Arabidopsis* mutant plants (lane co-IP, *phyB*) or in samples processed without anti-*phyB* antibody (lane -Ab, *cry2+*) indicated the specificity of the interaction. Reprobing the same blot with anti-*phyB* antibody (anti-*phyB* western blot) confirmed *phyB* immunoprecipitation (lane co-IP, *cry2+*) and the absence of signals in negative controls (lanes co-IP, *phyB* and *cry2+*, -Ab). These results are consistent with a direct, physical interaction between *phyB* and *cry2* in *Arabidopsis* plants.

Previous reports described the nuclear localization of *Arabidopsis cry2* (refs 14, 15). Both human and mouse *cry2* also localize in the nucleus³ suggesting that the nuclear localization is important for *cry2* function. We therefore examined the effects of light in *cry2* subcellular localization using a fusion protein of *cry2* with the red fluorescent protein (RFP). In transfected BY-2 protoplasts maintained in the dark, *cry2*-RFP appeared homogeneously distributed within the nucleus (Fig. 2a). Rapid scanning with the blue laser line

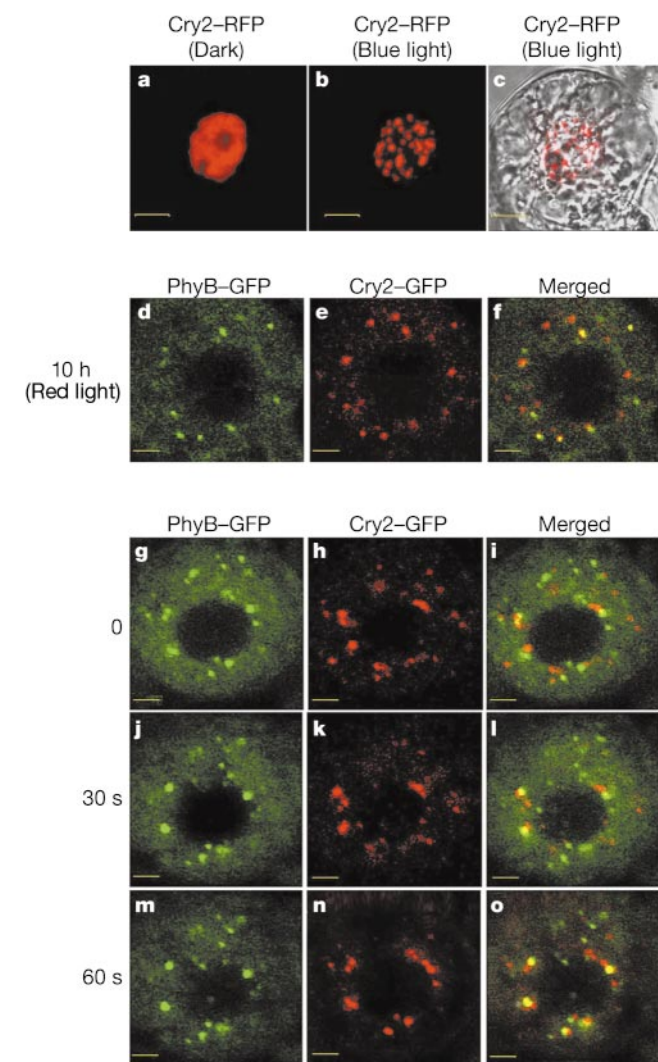


Figure 2 Effects of light on *cry2*-RFP subcellular distribution and colocalization with *phyB*-GFP in BY-2 cells. **a**, In the dark, *cry2*-RFP appeared distributed homogeneously in the nucleus. **b**, Formation of nuclear speckles after successive scanning with the blue laser of the confocal microscope. **c**, Superimposition of bright-field and fluorescent images. **d**, **e**, Nucleus of a cell co-expressing *phyB*-GFP (**d**) and *cry2*-RFP (**e**) examined after 10 h in continuous red light and scanning with the 488-nm laser line of the confocal microscope for *phyB* (**d**) and *cry2* (**e**) speckle formation. **f**, *phyB*-GFP and *cry2*-RFP colocalization after superimposing the images. **g**–**o**, *phyB*-GFP (**g**, **j**, **m**) and *cry2*-RFP (**h**, **k**, **n**) nuclear speckles examined at 30-s intervals in the same nucleus. Yellow in merged images (**i**, **l**, **o**) indicates sites where *phyB* and *cry2* colocalize. Scale bar, 5 μ m (**a**–**c**); 2 μ m (**d**–**o**).

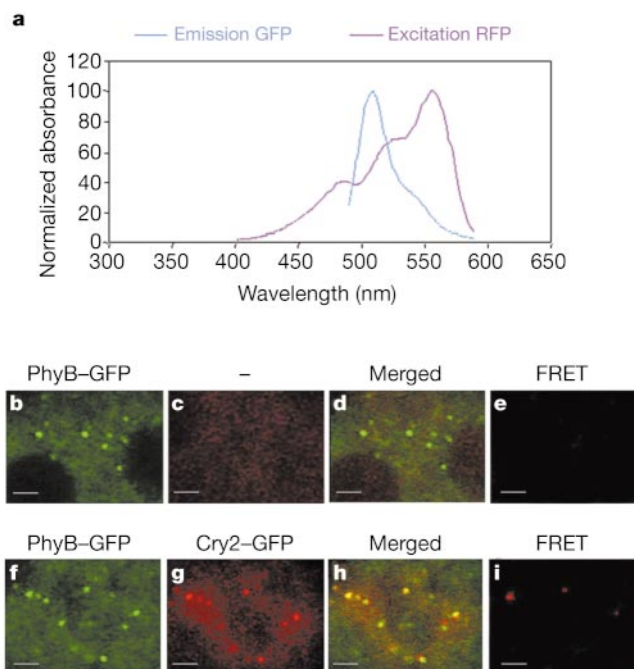


Figure 3 FRET analysis of *phyB*-GFP and *cry2*-RFP interaction. **a**, Spectral overlap (J_{da}) between the GFP emission and RFP excitation spectra. Using a value of 22,500 $M^{-1} cm^{-1}$ for the molar extinction coefficient (ϵ_a) of RFP, J_{da} was calculated as $1.67 \times 10^{-13} M^{-1} cm^{-1}$. From this value and a quantum yield (ϕ_d) of 0.60 for GFP, a refractive index (n) of 1.4 and a dipole orientation factor (κ^2) of 2/3, we calculated a critical transfer distance (R_0) of 5 nm. **b**, BY-2 cells transfected with *phyB*-GFP examined with the 488-nm laser of a confocal microscope using a FVX-BA550RIF filter. **c**, Background signal of the red channel examined with a 568-nm krypton laser and using an FVX-BA585IF filter. **e**, Adjustment of the threshold settings and background subtraction to eliminate crosstalk between the channels. **f**, **g**, **i**, Cells co-transfected with *phyB*-GFP (**f**) and *cry2*-RFP (**g**) excited with the 488-nm laser and observed using an FVX-BA585IF filter (**i**). **d**, **h**, Merged images of green and red channels. Scale bar, 2 μ m.

of the confocal microscope or irradiation with blue light induced the formation of nuclear speckles (Fig. 2b). Superimposition of bright-field and fluorescent images clearly revealed the nuclear accumulation of cry2 (Fig. 2c). Cry2-RFP did not induce speckles in cells maintained in the dark, indicating that light (not over-expressed proteins) may be responsible for the formation of these speckles. Cry2-GFP fusion protein also localized in nuclear speckles, indicating that these speckles were not due to artefactual expression of RFP in plants. Fluorescent speckles were never observed when protoplasts were transfected with non-fused RFP or RFP fused with cry1 (data not shown).

Previous studies reported that phyB translocates to nuclear

speckles under continuous red light^{16,17}. To examine the subcellular relationship between phyB and cry2, we co-transfected BY-2 protoplasts with phyB-GFP and cry2-RFP, and examined the fluorescent signal by confocal microscopy. Co-transfected protoplasts were maintained 8 h in the dark followed by irradiation with continuous red light for accumulation of phyB-GFP in nuclear speckles (Fig. 2d). These cells were then irradiated with blue light or scanned with the blue laser of the confocal microscope to induce cry2 speckles (Fig. 2e). Superimposing both images revealed that some of the speckles where cry2 and phyB accumulated perfectly colocalized in the nucleus (yellow stain in Fig. 2f). PhyB and cry2 colocalize in some but not all of these nuclear speckles. It is possible

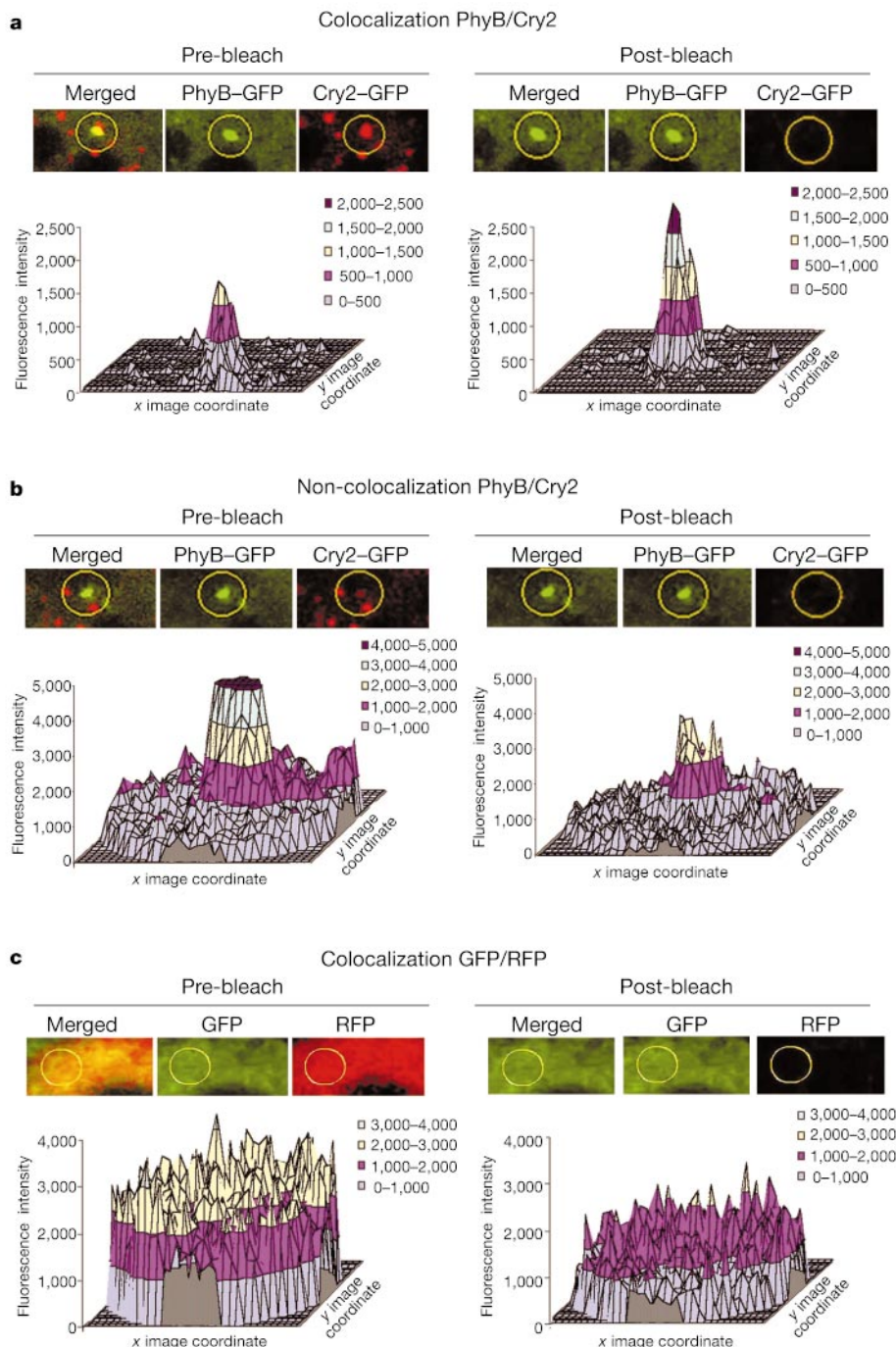


Figure 4 FRET microscopy by acceptor photobleaching. **a, b**, Fluorescence intensity of phyB-GFP in a demarcated region (yellow circle) was measured before and after cry2-RFP photobleaching in samples where phyB-GFP and cry2-RFP colocalized (**a**), or did

not colocalize (**b**). **c**, Fluorescent intensities of GFP before and after RFP photobleaching were also compared in samples where non-fused GFP and RFP colocalized. Results consistent with those shown here were obtained in five independent experiments.

that both photoreceptors might also have independent functions and that they recruit, in these speckles, other nuclear factors involved in light responses. We next investigated the changes of phyB–GFP and cry2–RFP distribution over time in protoplasts immobilized in agarose and examined by confocal microscopy at 30 s intervals. Figure 2g–i depicts an optical section of a nucleus showing phyB–GFP and cry2–RFP distribution in speckles that did not colocalize (row 0). Scanning at 30-s intervals the same cell revealed that some (row 30 s) or most (row 60 s) of cry2 speckles did colocalize with phyB (yellow stain in merged image). Together, these results demonstrate the light-dependent colocalization of phyB and cry2 in specific nuclear speckles.

We next used fluorescent resonance energy transfer (FRET) microscopy to determine whether the nuclear speckles are the sites where phyB and cry2 interact. FRET is a quantum mechanical process by which a fluorescent molecule, the donor, transfers energy by a non-radiative mechanism to an appropriately positioned chromophore, the acceptor¹⁸. Successful methods for the detection of intramolecular FRET signals in living plant cells have been developed using chromophore mutants of GFP¹⁹. We report, for the first time to our knowledge, the use of GFP–RFP combination for intermolecular detection of FRET signals in living cells. Before FRET analysis, we calculated GFP–RFP spectral overlap ($J_{da} = 1.67 \times 10^{-13} \text{ M}^{-1} \text{ cm}^3$) and Förster distance ($R_0 = 5 \text{ nm}$) to show that GFP–RFP combination is a suitable donor–acceptor pair for FRET (Fig. 3a). The rationale of our experiments resides in the fact that FRET would be detected only when the fluorophores are directly interacting and not simply colocalizing. We first calibrated the microscope system using cells transfected only with phyB–GFP (Fig. 3b) and no cry2–RFP (Fig. 3c). The samples were used to subtract the background and to adjust appropriately the threshold settings (Fig. 3e). This control verified that the signals obtained in FRET were not due to channel overlap. To measure FRET, we excited cells co-expressing phyB–GFP (Fig. 3f) and cry2–RFP (Fig. 3g) with the 488 nm laser-line, and observed cry2–RFP emission using a FVX-BA585IF filter. Significant FRET emission of cry2–RFP (Fig. 3i, FRET) was observed in speckles that colocalized with phyB–GFP (see yellow in merged image, Fig. 3h). These results indicate the direct molecular interaction of phyB and cry2 in nuclear speckles and confirm the previous results obtained in co-immunoprecipitation experiments.

FRET measurement as described above can lead to errors, mainly owing to differences in the concentration of phyB–GFP and cry2–RFP or to the direct excitation of RFP at the GFP excitation wavelength. Furthermore, background subtraction to minimum fluorescence values (Fig. 3e) might mask less strong FRET signals and eliminate the possibility of obtaining FRET efficiency maps. We therefore evaluated FRET using a different method that is based on the increase of the donor fluorescence after photochemical destruction of the acceptor²⁰. Cells co-expressing phyB–GFP and cry2–RFP (Fig. 4a) were maintained in continuous red light and imaged, first with the 488-nm laser line (phyB–GFP) and later with the 568-nm laser line (cry2–RFP). We chose one of the speckles where phyB–GFP (green) and cry2–RFP (red) colocalized (yellow in merged image) and measured the fluorescence intensity of phyB–GFP in a region roughly 4 μm in diameter (demarcated by a yellow circle; Fig. 4a, pre-bleach). Cry2–RFP was subsequently photobleached by repeated scanning with the 568-nm laser line until no signal was detected (cry2–RFP, Fig. 4a post-bleach). We then measured the fluorescent intensity of phyB–GFP and compared it with that obtained in the same region before bleaching. A clear increase of phyB–GFP fluorescence was observed (compare pre- and post-bleach in Fig. 4a). This effect was not detected in phyB–GFP speckles that did not colocalize with cry2–RFP (Fig. 4b). In those cases, the levels of fluorescence intensity of phyB–GFP after cry2–RFP bleaching remained the same or slightly decreased as compared with those observed before bleaching (pre- and post-bleach in

Fig. 4b). To verify that FRET signals were not due to fluorophore colocalization, we carried out an additional negative control with cells co-expressing non-fused RFP and non-fused GFP constructs. Bleaching RFP that colocalized with GFP did not increase GFP fluorescent intensity relative to its fluorescence before bleaching (Fig. 4c, pre- and post-bleach). Our results show that phyB–GFP fluorescence increased at sites where FRET was occurring, confirming the physical interaction between cry2 and phyB in nuclear speckles.

The light-induced nuclear compartmentalization of phytochromes^{16,17}, the discovery that they possess kinase activity²¹ and the identification of some of their interacting partners^{22–24} have provided crucial understanding of how phytochromes transmit light signals. These findings suggest that regulation of protein expression, phosphorylation and subcellular compartmentalization may have key roles in mediating light-induced physiological responses in plants. Cry1 and cry2 are phosphorylated by the kinase activity of phyA²⁵. Our results show that phyB and cry2 interact in specific nuclear speckles. It is possible that these speckles act as ‘transcriptosomes’ similarly to some factors described in animal cells that accumulate in a characteristic speckled pattern and are involved in processing and transcription of RNA²⁶. Light-induced responses are also mediated by alteration of protein stability. Cry2 is strongly downregulated by blue-light¹³, so the nuclear speckles could represent sites where cry2 is degraded. Recent studies have suggested that novel clock-associated components contain domains involved in targeting specific substrates for proteolytic degradation^{27,28}. Further experiments need to clarify whether these clock components are involved in cry2 degradation, and whether phyB has a role in this process. □

Methods

Measurements of growth and flowering

Wild-type and *cry2* seeds were grown for 6 d in constant white light ($60 \mu\text{mol m}^{-2} \text{ s}^{-1}$) at 22 °C before being either maintained in constant white light (R/FR ratio 8.53) or transferred to white light supplemented with FR (R/FR ratio 0.05). A low R/FR ratio light was generated by supplementing white light with FR light from two LED light sources (Quantum devices, Barneveld, WI). Light measurements were made using an LI 1800/12 spectroradiometer (Li-Cor, Lincoln, NE). Flowering time was measured as the number of rosette leaves at bolting (production of a 1-cm high cauline stem). Data represent the mean ($\pm$ s.e.) of 6–8 plants for each treatment. Hypocotyl length was measured after 10 d in the treatments using Scion Image software. Data represent the mean ($\pm$ s.e.) of 28–31 seedlings for each treatment.

Light input to the circadian clock

Wild-type and *cry2* seeds expressing a *CAB2::LUCIFERASE* reporter construct⁸ were grown in 12 h white light/12 h dark cycles for 6 d before being transferred to constant red light (600–700 nm), blue light (400–500 nm) or white light at the intensities indicated. The rhythm of bioluminescence, representing *CAB2* transcription, was followed as described⁷ and period length was calculated by fitting a cosine wave function to the time series for each seedling⁷. Data represent the mean ($\pm$ s.e.) of 5–18 seedlings, representative of 2–4 independent experiments.

Coimmunoprecipitation experiments

Cry2 overexpression plants¹³ (provided by C. Lin, Univ. California, Los Angeles) were grown in 12 h white light/12 h dark cycles and processed for co-immunoprecipitation in the first 15 min of the light cycle. PhyB immunoprecipitation was performed at 4 °C for 2 h using anti-phyB antibody (provided by A. Nagatani, Kyoto Univ., Kyoto) in a buffer containing 10 mM HEPES pH 7.5, 10% glycerol, 100 mM NaCl, 0.2% Triton X-100, 1 mM EDTA and 1 mM dithiothreitol with a mixture of protease inhibitors (Sigma). Gamma-bind plus sepharose beads (Pharmacia) were used to precipitate the immunoprotein complexes. Cry2 detection was performed using anti-cry2 antibody (provided by C. Lin, Univ. California, Los Angeles).

Plasmid construction and confocal imaging

Red fluorescent protein (RFP) coding region was obtained from the vector pDsRed1-N1 (Clontech Laboratories). Cry2 coding sequence was obtained from C. Lin. The fusion protein cry2–RFP was cloned into the vector pRTL2 which contains a CaMV 35S promoter²⁹. Colocalization of phyB–GFP¹⁶ and cry2–RFP was examined in BY-2 protoplasts prepared and transfected as described³⁰. Protoplasts were maintained 8 h in the dark followed by irradiation with continuous red light ($15 \mu\text{mol m}^{-2} \text{ s}^{-1}$) for 10 h. Cry2 speckle formation was induced after irradiation with blue light ($5 \mu\text{mol m}^{-2} \text{ s}^{-1}$) for 1 min or after successive scanning with the 488-nm laser of a confocal laser-scanning microscope (IX70; Olympus Corp). To examine GFP fluorescence, we scanned samples with a 488-nm argon

laser line using a 550-nm (FVX-BA550RIF) barrierfilter and $\times 60$ 1.4 NA PlanApo oil-immersion objective. RFP signal was examined with a 568-nm krypton laser using an FVX-BA585IF filter.

FRET analysis

Normalized fluorescence emission spectrum of GFP (F_d) (provided by G. Patterson and D. Piston, Vanderbilt Univ., Nashville, Tennessee) and absorption spectrum of DsRFP (provided by Clontech) were used to calculate GFP–RFP spectral overlap (J_{da}) and Förster distance (R_0) according to the equations¹⁸:

$$J_{da} = \frac{\int_0^\infty F_d(\lambda)\epsilon(\lambda)\lambda^4 d\lambda}{\int_0^\infty F_d(\lambda)d\lambda} \quad (1)$$

and

$$R_0^6 = (8.75 \times 10^{-25}) \kappa^2 n^{-4} \phi_d J_{da} \quad (2)$$

where ϵ is the molar extinction coefficient of RFP; ϕ_d is the quantum yield for GFP; n is the refractive index; and κ^2 is the dipole orientation factor. FRET signals were examined by excitation with the 488-nm argon laser and subsequent detection of cry2–RFP using a FVX-BA585IF filter. Single fluorophore imaging allowed adjustments of the threshold settings and background subtractions to reduce the levels of crosstalk between the channels. To measure FRET signals by acceptor photobleaching, a pre-bleached phyB–GFP image was acquired, and the fluorescent intensity in a region of interest of 4 μ m in diameter was recorded using the software provided by the manufacturer (Fluoview). Cry2–RFP fluorescence was then photodestroyed by repeated scanning with the 568-nm laser line. A second post-bleach image of phyB–GFP in the same region of interest was acquired. After corrections for changes in image registration²⁰, the fluorescent intensities of the two phyB–GFP images (pre- and post-bleached) were compared. PhyB–GFP speckles that did not colocalize with cry2–RFP and colocalization of non-fused GFP and RFP were similarly processed.

Received 14 June; accepted 29 August 2000.

- Kendric, R. E. & Kronenberg, G. H. M. *Photomorphogenesis in Plants* (Kluwer Academic, Dordrecht, 1994).
- Quail, P. H. *et al.* Phytochromes: photosensory perception and signal transduction. *Science* **268**, 675–680 (1995).
- Cashmore, A. R., Jarillo, J. A., Wu, Y. & Liu, D. Cryptochromes: Blue light receptors for plants and animals. *Science* **284**, 760–765 (1999).
- Chory, J. Light modulation of vegetative development. *Plant Cell* **9**, 1225–1234 (1997).
- Briggs, W. R. & Huala, E. Blue-light photoreceptors in higher plants. *Annu. Rev. Cell Dev. Biol.* **15**, 33–62 (1999).
- Casal, J. J. Phytochromes, cryptochromes, phototropin: photoreceptor interactions in plants. *Photochem. Photobiol.* **71**, 1–11 (2000).
- Somers, D. E., Devlin, P. F. & Kay, S. A. Phytochromes and cryptochromes in the entrainment of the *Arabidopsis* circadian clock. *Science* **282**, 1488–1490 (1998).
- Millar, A. J., Straume, M., Chory, J., Chua, N. H. & Kay, S. A. The regulation of circadian period by phototransduction pathways in *Arabidopsis*. *Science* **267**, 1163–1166 (1995).
- Whitelam, G. C. & Devlin, P. F. Roles of different phytochromes in *Arabidopsis* photomorphogenesis. *Plant Cell Environ.* **20**, 752–758 (1998).
- Guo, H., Yang, H., Mockler, T. C. & Lin, C. Regulation of flowering time by *Arabidopsis* photoreceptors. *Science* **279**, 1360–1363 (1998).
- Mockler, T. C., Guo, H., Yang, H., Duong, H. & Lin, C. Antagonistic actions of *Arabidopsis* cryptochromes and phytochrome B in the regulation of floral induction. *Development* **126**, 2073–2082 (1999).
- Reed, J. W., Nagpal, P., Poole, D. S., Furuya, M. & Chory, J. Mutations in the gene for the red/far red light receptor phytochrome B alter cell elongation and physiological responses throughout *Arabidopsis* development. *Plant Cell* **5**, 147–157 (1993).
- Lin, C. *et al.* Enhancement of blue-light sensitivity of *Arabidopsis* seedlings by blue light receptor cryptochrome 2. *Proc. Natl Acad. Sci. USA* **95**, 2686–2690 (1998).
- Guo, H., Duong, H., Ma, N. & Lin, C. The *Arabidopsis* blue light receptor cryptochrome 2 is a nuclear protein regulated by a blue light-dependent post-transcriptional mechanism. *Plant J.* **19**, 279–287 (1999).
- Kleiner, O., Kircher, S., Harter, K. & Batschauer, A. Nuclear localization of the *Arabidopsis* blue light receptor cryptochrome 2. *Plant J.* **19**, 289–296 (1999).
- Yamaguchi, R., Nakamura, M., Mochizuki, N., Kay, S. A. & Nagatani, A. Light-dependent translocation of a phytochrome B–GFP fusion protein to the nucleus in transgenic *Arabidopsis*. *J. Cell Biol.* **145**, 437–445 (1999).
- Kircher, S. *et al.* Light quality-dependent nuclear import of the plant photoreceptors phytochrome A and B. *Plant Cell* **11**, 1445–1456 (1999).
- Wu, P. & Brand, L. Resonance energy transfer: methods and applications. *Anal. Biochem.* **218**, 1–13 (1994).
- Gadella, T. W. J., van der Krogt, G. N. M. & Bisseling, T. GFP-based FRET microscopy in living plant cells. *Trends Plant Sci.* **4**, 287–291 (1999).
- Bastiaens, P. I. H. & Jovin, T. M. in *Cell Biology: A Laboratory Handbook* (ed. Celis, J. E.) 136–146 (Academic, New York, 1998).
- Yeh, K. C. & Lagarias, J. C. Eukaryotic phytochromes: light-regulated serine/threonine protein kinases with histidine kinase ancestry. *Proc. Natl Acad. Sci. USA* **95**, 13976–13981 (1998).
- Ni, M., Tepperman, J. M. & Quail, P. H. PIF3, a phytochrome-interacting factor necessary for normal photoinduced signal transduction, is a novel basic helix–loop–helix protein. *Cell* **95**, 657–667 (1998).
- Fankhauser, C. *et al.* PKS1, a substrate phosphorylated by phytochrome that modulates light signaling in *Arabidopsis*. *Science* **284**, 1539–1541 (1999).
- Choi, G. *et al.* Phytochrome signalling is mediated through nucleoside diphosphate kinase 2. *Nature* **401**, 610–613 (1999).
- Ahmad, M., Jarrillo, J. A., Smirnova, O. & Cashmore, A. R. The CRY1 blue light photoreceptor of *Arabidopsis* interacts with phytochrome *in vitro*. *Mol. Cell* **1**, 939–948 (1998).

- Lamond, A. I. & Earnshaw, W. C. Structure and function in the nucleus. *Science* **280**, 547–553 (1998).
- Somers, D. E., Schultz, T. F., Milnamow, M. & Kay, S. A. ZEITLUPE encodes a novel clock-associated PAS protein from *Arabidopsis*. *Cell* **101**, 319–329 (2000).
- Nelson, D. C., Lasswell, J., Rogg, L. E., Cohen, M. A. & Bartel, B. FKF1, a clock-controlled gene that regulates the transition to flowering in *Arabidopsis*. *Cell* **101**, 331–340 (2000).
- Carrington, J. C., Freed, D. D. & Leinicke, A. J. Bipartite signal sequence mediates nuclear translocation of the plant potyviral N1a protein. *Plant Cell* **3**, 953–962 (1991).
- Más, P. & Beachy, R. N. Replication of tobacco mosaic virus on endoplasmic reticulum and role of the cytoskeleton and virus movement protein in intracellular distribution of viral RNA. *J. Cell Biol.* **147**, 945–958 (1999).

Acknowledgements

We thank S. L. Harmer, T. F. Schultz and M. J. Yanovsky for critical comments of the manuscript. We are grateful to C. Lin for providing cry2 cDNA, the cry2 overexpression line and cry2 antibody. We also thank A. Nagatani for the phyB–GFP construct and anti-phyB antibody. We thank G. Patterson and D. Piston for GFP spectra data, and Clontech for DsRFP spectra data. We are grateful to D. Millar and T. K. Nomanbhoy for helpful discussion on FRET analysis. Research support came from the NIH. P.M. was supported by a Novartis Agricultural Discovery Institute and P.F.D. was supported by an European Molecular Biology Organization long-term fellowship and the NSF. S.P. was supported by TSRI graduate program.

Correspondence and requests for materials should be addressed to S.A.K. (e-mail: stevek@scripps.edu).

Intracellular action of the cytokine MIF to modulate AP-1 activity and the cell cycle through Jab1

Robert Kleemann*, Angelika Hausser†, Georg Geiger*, Ralf Mischke*, Anke Burger-Kentischer*, Oliver Flieger*, Franz-Josef Johannes†, Thierry Roger§, Thierry Calandra§, Aphrodite Kapurniotou||, Matthias Grell†, Doris Finkelmeier*, Herwig Brunner* & Jürgen Bernhagen*

* Laboratory of Biochemistry/Institute for Interfacial Engineering and †Institute for Cell Biology and Immunology, University of Stuttgart, Nobelstrasse 12, D-70569 Stuttgart, Germany

‡ Fraunhofer IGB, Nobelstrasse 12, D-70569 Stuttgart, Germany

§ Division of Infectious Diseases, Centre Hospitalier Universitaire Vaudois, CH-1011 Lausanne, Switzerland

|| Physiological-chemical Institute, University of Tübingen, Hoppe-Seyler-Strasse 4, D-72076 Tübingen, Germany

Cytokines are multifunctional mediators that classically modulate immune activity by receptor-mediated pathways. Macrophage migration inhibitory factor (MIF) is a cytokine that has a critical role in several inflammatory conditions^{1–3} but that also has endocrine^{4,5} and enzymatic functions^{6,7}. The molecular targets of MIF action have so far remained unclear. Here we show that MIF specifically interacts with an intracellular protein, Jab1, which is a coactivator of AP-1 transcription^{8,9} that also promotes degradation of the cyclin-dependent kinase inhibitor p27^{Kip1} (ref. 10). MIF colocalizes with Jab1 in the cytosol, and both endogenous and exogenously added MIF following endocytosis bind Jab1. MIF inhibits Jab1- and stimulus-enhanced AP-1 activity, but does not interfere with the induction of the transcription factor NFκB. Jab1 activates c-Jun amino-terminal kinase (JNK) activity and enhances endogenous phospho-c-Jun levels, and MIF inhibits these effects. MIF also antagonizes Jab1-dependent cell-cycle regulation by increasing p27^{Kip1} expression through stabilization of p27^{Kip1} protein. Consequently, Jab1-mediated rescue of fibroblasts from growth arrest is blocked by MIF. Amino acids 50–65

§Present address: Merck KGaA, Preclinical Research Oncology, D-64271 Darmstadt, Germany.

and Cys 60 of MIF are important for Jab1 binding and modulation. We conclude that MIF may act broadly to negatively regulate Jab1-controlled pathways and that the MIF–Jab1 interaction may provide a molecular basis for key activities of MIF.

There is evidence for intracellular functions of MIF^{1,4,6,7}, and we thought that these intracellular functions might be based on direct interactions of MIF with intracellular proteins. To identify a putative MIF-binding protein, we carried out a yeast two-hybrid screen¹¹. As MIF is abundantly expressed in the brain¹², we used a human brain complementary DNA library and full-length MIF cDNA¹³ fused to the DNA-binding domain of GAL4 as bait. Out of 1.1×10^7 transfectants, three types of clone interacted specifically when tested for nutritional selection and β -galactosidase activity (see Supplementary Information). One clone contained a cDNA insert with almost the entire coding sequence (amino acids 20–335) of human Jab1 (refs 8–10, 14, 15). The two other types of clone were identified as human MIF, confirming the oligomerization of MIF¹⁴.

We verified a specific association between MIF and Jab1 by co-immunoprecipitation and pull-down experiments of MIF–Jab1 complexes. First, we analysed the formation of MIF–Jab1 complexes *in vitro*. When added to Jab1-programmed reticulocyte lysates, purified soluble recombinant MIF (rMIF)¹³ specifically bound to *in vitro* translated biotin-labelled full-length Jab1 (Fig. 1a). We also observed specific complex formation *in vitro* between MIF and Jab1 when both proteins were expressed in the *in vitro* translation system. Complexes of biotin-labelled MIF–EGFP (enhanced green fluorescent protein) fusion protein and non-biotinylated Jab1 were specifically precipitated (Fig. 1b). Furthermore, biotinylated rMIF bound to endogenous Jab1 *in vivo* when it was added to Raw 264.7

macrophages or HeLa cells (Fig. 2a, inset). This last finding also shows that MIF that has been added exogenously to cells can bind to endogenous intracellular Jab1 following endocytosis. We also investigated the *in vivo* interaction of the endogenous partners expressed at normal levels. Endogenous Jab1–MIF complexes were precipitated from untreated 293T cells. Specific coprecipitation of MIF was detected by anti-MIF western blot when anti-Jab1 antibody but not control antibody was used for coprecipitation (Fig. 1c). Together, the two-hybrid screen and the various coprecipitation data of the overexpressed and endogenous binding partners (see also Supplementary Information) indicate that MIF specifically binds to Jab1 both *in vivo* and *in vitro*.

To confirm that MIF can interact with intracellular Jab1 (refs 8–10, 17), we analysed the localization of endogenously expressed MIF; and, as extracellular MIF is critical for the immunological functions of MIF¹, we also tested whether extracellular MIF is targeted to the cytosol. Detection of endogenous MIF with anti-MIF antibody, and Cy-2 staining or transient expression of MIF–EGFP fusion protein in HeLa cells and comparison with compartment-specific marker proteins revealed that both endogenous MIF

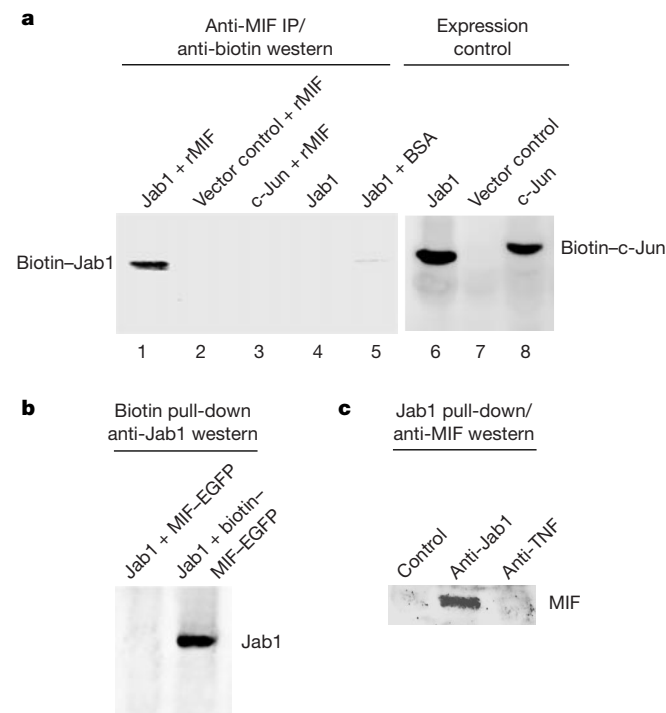


Figure 1 MIF specifically interacts with Jab1. **a**, Interaction *in vitro*. Left, complexes of rMIF and biotin–Jab1 expressed in TNT lysates were precipitated by MIF-specific antibody, and biotinylated proteins were detected by western blot. Right, expression control. **b**, Interaction of immobilized MIF with Jab1. TNT-translated biotin–MIF–EGFP immobilized on streptavidin beads or unlabelled control was incubated with TNT-translated Jab1, complexes were isolated, and Jab1 was detected by immunoblotting. **c**, Interaction of MIF and Jab1 *in vivo*. Endogenous Jab1–MIF complexes were immunoprecipitated from 293T cell lysates with anti-Jab1 antibody and MIF was detected by immunoblotting. Precipitation with anti-TNF antibody or beads alone served as controls.

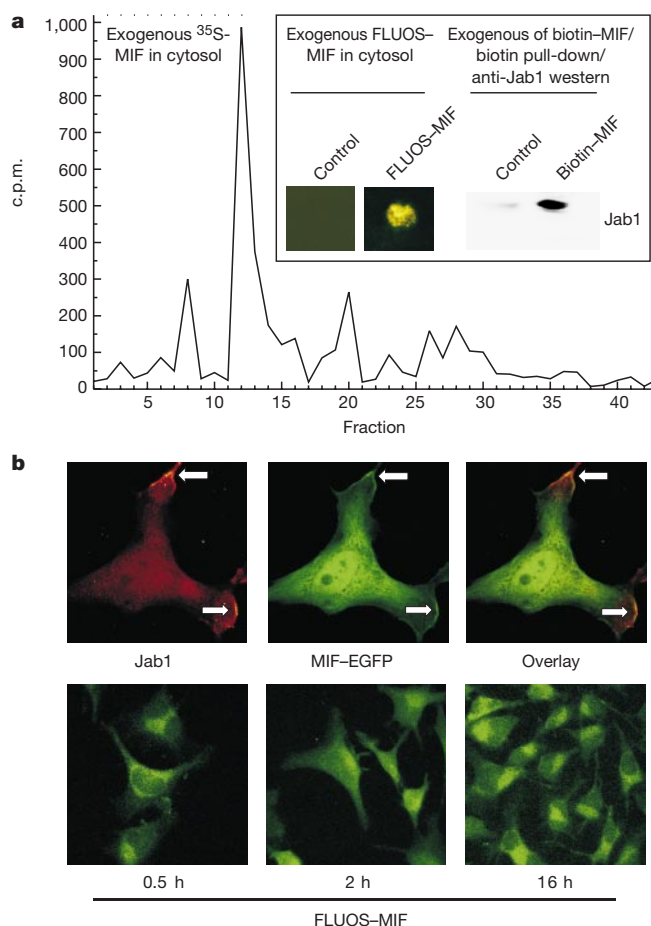


Figure 2 MIF is targeted to the cytosol and colocalizes with Jab1. **a**, Cytosolic targeting of exogenous rMIF and binding to Jab1. Gel-filtration chromatogram of cytosolic Raw cell fractions after incubation of cells with [³⁵S]rMIF. For elution of labelled and endocytosed controls see Supplementary Information. Left inset, fluorescence microscopy of HeLa cells after endocytosis of FLUOS–MIF. Right inset, cytosolic targeting of biotin–MIF and binding to Jab1. MIF–Jab1 complexes from Raw cells after incubation with biotin–MIF were precipitated with streptavidin beads and immunoblots were developed for Jab1. **b**, MIF colocalizes with Jab1 in the cytosol of HeLa cells (confocal microscopy). Top row, colocalization of overexpressed MIF–EGFP (green) with endogenous Jab1 (red). Arrows indicate stained areas near the peripheral plasma membrane. Bottom row, kinetics of cytosolic targeting of exogenous FLUOS–MIF.

and endogenously overexpressed MIF are predominantly located in the cytosol (see Supplementary Information). $[^{35}\text{S}]\text{rMIF}$ (Fig. 2a) and fluorescently labelled rMIF (FLUOS-MIF) (Fig. 2a, left inset) exogenously added to Raw or HeLa cells were detected in the cytosol of target cells after cellular exposition and removal of surface-bound label. Of note, endocytosis of rMIF was not saturable over a wide concentration range of rMIF and was not competitively inhibited by an excess of cold rMIF (see Supplementary Information). Notably, endocytosis of biotin-MIF with subsequent intracellular binding to endogenous Jab1 was observed (Fig. 2a, right inset). Confocal microscopic analysis in HeLa cells of endogenously overexpressed MIF-EGFP or added FLUOS-MIF confirmed the cytosolic localization/targeting of MIF and directly showed colocalization of MIF with Jab1 in the cytosol. An enrichment of Jab1 has been found near the peripheral plasma membrane⁹, which agrees with the distinct

MIF-Jab1 costaining that we observed in this area (Fig. 2b). Together, these data show that exogenous MIF can interact with Jab1 after uptake into target cells, and suggest that cytokine activities of MIF can be mediated by Jab1-dependent pathways.

The transcriptional coactivator function of Jab1 is associated with the enhancement of AP-1 transcriptional activity⁸. We tested whether MIF, by binding to Jab1, could modulate this AP-1 activity. We measured the AP-1-dependent reporter gene activity in 293T cells transiently expressing the collagenase TRE luciferase reporter. Recombinant MIF, in a concentration-dependent manner, inhibited potentiation of AP-1 reporter gene activity that was induced by cotransfected Jab1, with almost complete inhibition observed at 1 μM MIF (Fig. 3a). Inhibition of AP-1 activity by MIF was not a secondary effect of MIF-mediated altered cell growth (see Supplementary Information). Analysis of DNA-binding activity by electrophoretic mobility-shift assay (EMSA) of nuclear extracts from 293T cells transiently overexpressing Jab1 showed that Jab1-mediated enhancement of c-Jun binding to the AP-1 probe was reduced threefold after co-incubation with rMIF (data not shown). This confirmed the direct involvement of the MIF-Jab1 interaction in MIF-mediated inhibition of AP-1 activity. MIF, in a concentration-dependent manner, also inhibited tumour necrosis factor (TNF)- and ultraviolet-stress-induced AP-1 transcriptional activity (Fig. 3b; and Supplementary Information). We can confirm the specificity of MIF signalling activity for Jab1-dependent pathways as neither exogenously added nor endogenous MIF interfered with NF κ B activity. Recombinant MIF (1 μM) had no effect on TNF-induced NF κ B reporter gene activity in 293T cells (see Supplementary Information), and reduction of the endogenous MIF content by

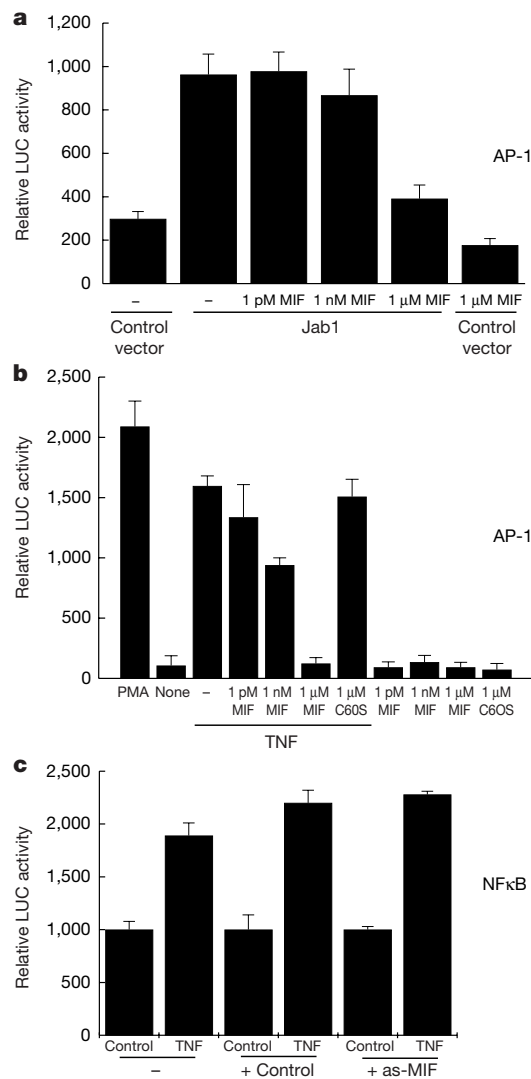


Figure 3 MIF inhibits stimulated AP-1 activity. **a**, MIF inhibits Jab1-mediated activation of AP-1 reporter gene activity in 293T cells. Transfections with pCI-neo-Jab1 or control vector are compared with or without (–) rMIF. **b**, MIF inhibits TNF-induced AP-1 activity. Same as **a**, but with TNF instead of Jab1 induction. Stimulation by PMA was used as a control. For transfection efficiencies see Supplementary Information. Data represent mean $\pm$ s.d. of four determinations, and are representative of three experiments. C60S, mutant C60S MIF. **c**, MIF does not inhibit stimulated NF κ B reporter gene activity. TNF-induced NF κ B from antisense MIF macrophages with reduced content of endogenous MIF (+ as-MIF) is compared with activities of control (+ control) and non-transfected (–) cells. The mean $\pm$ s.d. of three measurements is given. LUC, relative luciferase activity.

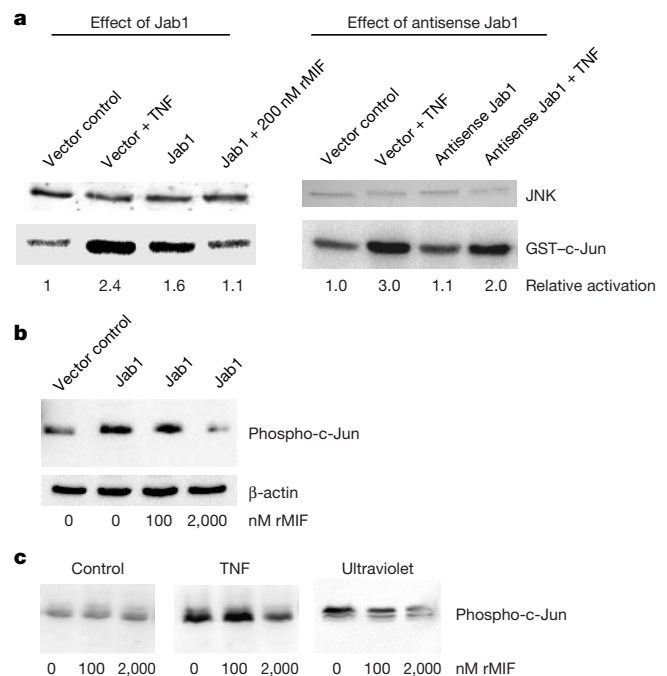


Figure 4 Jab1 enhances JNK activity and phospho-c-Jun levels and MIF inhibits these effects. **a**, rMIF inhibits Jab1-mediated enhancement of JNK activity. Left, comparison of phosphorylation of GST-c-Jun(1–79) by JNK immunoprecipitates from 293T cells transfected with pCI-neo-Jab1 versus empty vector and treated with rMIF or buffer. Relative activation indicates band intensities; control is TNF induction. Right, as left but cotransfection of antisense-Jab1 plasmid where indicated. Control, immunoblots of JNK levels; transfection efficiencies, 40–50%; variances, $\pm$ 10%; induction of JNK by Jab1, 1.6-fold $\pm$ s.d. of 0.08; $n = 3$. **b**, Inhibition by MIF of Jab1-induced phosphorylation of endogenous c-Jun. Incubations were as in **a**, but immunoblots of phospho-c-Jun were analysed. β -Actin immunoblots served as a control. **c**, As **b**, but phosphorylation of c-Jun induced by TNF or ultraviolet stress.

an antisense approach in macrophages did not affect TNF-induced NF κ B reporter gene activity (Fig. 3c).

Cytokines stimulate AP-1 activity through several pathways¹⁸, and we therefore determined whether MIF might regulate AP-1 activity by additional mechanisms, that is, by effects on JNK activity. Transient overexpression of Jab1 in 293T cells revealed that JNK immunoprecipitates phosphorylated glutathione S-transferase (GST)-c-Jun(1-79) more potently than immunoprecipitates from control cells (Fig. 4a, left). Involvement of Jab1 in JNK activation was confirmed by an observed partial blockage of endogenous Jab1 by cotransfection of an antisense Jab1 construct (Fig. 4a, right). We next investigated whether MIF would inhibit Jab1-stimulated JNK activity. Recombinant MIF reversed enhancement of JNK activity by Jab1, with ~200 nM MIF conferring complete inhibition (Fig. 4a, left). Similarly, endogenously overexpressed MIF-EGFP fully suppressed Jab1-induced stimulation of JNK activity (data not shown), showing that both endogenously expressed and exogenously added MIF can counter-regulate Jab1 action. Moreover, MIF inhibited (about 2.5-fold) TNF-induced JNK activity (see Supplementary Information). The precise mechanism of the effect of MIF on the JNK pathway is currently unclear, as neither MIF nor Jab1 measurably bound to JNK directly; Jab1 did not enhance JNK, phospho-JNK or ultraviolet-induced phospho-JNK levels, excluding the effects of the Jab1-MIF pair on expression or phosphorylation of the JNK protein; and MIF did not affect Jab1-enhanced binding of JNK to immobilized GST-c-Jun(1-79) (data not shown; and see Supplementary Information). However, the *in vivo* relevance of the inhibitory effect of MIF on the

JNK pathway was revealed by our finding that treatment of cells with rMIF led to a marked reduction of endogenous phospho-c-Jun levels that had been induced by transient overexpression of Jab1 (Fig. 4b), or by agents such as TNF, ultraviolet (Fig. 4c) or anisomycin (data not shown). Overall, the AP-1 and JNK data indicated that the MIF-Jab1 pair could represent a critical counter-regulatory system in the modulation of AP-1 pathways.

To investigate whether MIF could also negatively regulate Jab1 action with regard to non-AP-1-related activities, we studied the effect of MIF on the cell-cycle inhibitor p27^{Kip1}, which binds to Jab1 and whose degradation is instigated by Jab1 (ref. 10). Contrary to Jab1, which suppresses p27^{Kip1} levels, MIF induced p27^{Kip1} levels in a dose-dependent manner, as indicated by immunoblots prepared from lysates of proliferating NIH 3T3 fibroblasts (Fig. 5a) and Jurkat T cells (not shown). This induction was Jab1-dependent, as p27^{Kip1} levels did not rise in response to rMIF when cells were transfected with antisense Jab1 construct (Fig. 5b), which almost completely suppressed endogenous Jab1 protein. p27^{Kip1} induces G1 growth arrest, and Jab1 can rescue serum-starved fibroblasts from growth arrest¹⁰. We found that overexpressed MIF-EGFP (Fig. 5c) or exogenous rMIF (see Supplementary Information) inhibited in a concentration-dependent manner both Jab1-induced reduction of serum dependence of fibroblasts and growth of proliferating fibroblasts. As MIF did not directly bind to p27^{Kip1}, and did not stimulate p27^{Kip1} mRNA or protein synthesis (see Supplementary Information), and as Jab1 promotes the proteasome-dependent degradation of p27^{Kip1} (ref. 10), we next analysed effects of MIF in the context of p27^{Kip1} degradation. When measured by immunoprecipitation from synchronized pulse-chase-labelled fibroblasts, p27^{Kip1} levels were higher in the presence of added rMIF (Fig. 5d, left). The stabilizing effect of MIF on p27^{Kip1} levels was not further enhanced by addition of the proteasome inhibitor LLnL (Fig. 5d, right). By contrast, enhancement of p27^{Kip1} levels by MIF alone was even more pronounced than that by LLnL alone. At the same time, addition of MIF slightly increased p27^{Kip1} levels in the presence of a protein synthesis blocker. As co-immunoprecipitation studies indicated that MIF partially interfered with p27^{Kip1}-Jab1 complex formation (see Supplementary Information), these data collectively suggest that MIF-mediated effects mirror p27^{Kip1}-mediated growth arrest and that this occurs through inhibition of Jab1-dependent degradation of p27^{Kip1}.

Of various rationally designed mutants of MIF investigated to date, mutant C60SMIF exhibits a clear structure activity profile. This mutant can be readily folded for use in activity assays but is devoid of the enzymatic oxido-reductase and immunological activity of MIF^{7,19}. We found that C60SMIF, while capable of binding to biotin-Jab1, did not inhibit TNF-induced AP-1 activity, did not reduce Jab1-induced enhancement of c-Jun DNA binding, and exhibited reduced p27^{Kip1}-inducing properties. Of note, a 16-residue MIF peptide spanning Cys 60, MIF(50-65), and the double mutant peptide Ser 57/Ser 60/MIF(50-65) strongly competed with wild-type MIF for Jab1 binding (Fig. 3b; Supplementary Information; and data not shown), indicating that this region is involved in the binding and modulation of Jab1.

The molecular basis of MIF action was previously unknown. To our knowledge, Jab1 is the first protein identified to interact with MIF, showing that a cytokine can modulate signalling pathways by intracellular interaction with a transcriptional coactivator. Our data further argue that signalling through MIF-Jab1 is independent of a potential MIF receptor. Notably, colocalization of MIF and Jab1 in areas near the peripheral plasma membrane suggests that there may be a connection between the MIF and integrin signalling pathways⁹. Antagonism of MIF on Jab1 appears to be a broader principal of transcriptional and cell-cycle control, as several Jab1-modulated functions such as AP-1 and JNK activation, p27^{Kip1} expression and cell growth are inhibited by MIF. Although the mechanism for the MIF-Jab1 antagonism remains to be explored in detail, our data

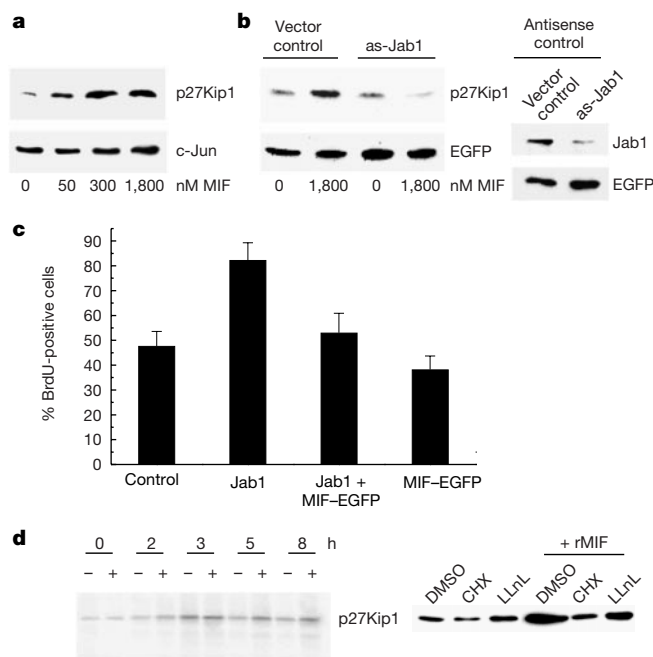


Figure 5 MIF stabilizes p27^{Kip1} protein and inhibits fibroblast growth in a Jab1-dependent manner. **a**, MIF induces p27^{Kip1} expression in fibroblasts (p27^{Kip1} immunoblots and c-Jun control blots). **b**, p27^{Kip1} induction by MIF is Jab1-dependent. Left, as **a**, but additional transfection of antisense Jab1 or control plasmid (efficiency control, pEGFP co-transfection). Right, the antisense construct reduces Jab1 in fibroblasts. **c**, Inhibition by MIF of Jab1-mediated reduction of serum dependence of fibroblasts. Proliferation of GFP-expressing cells is analysed in Jab1- and MIF-EGFP-overexpressing versus control vector-treated cells through 5-bromodeoxyuridine (BrdU) incorporation. Data are means $\pm$ s.d. of four determinations. **d**, MIF inhibits proteasome-mediated degradation of p27^{Kip1}. Left, MIF reduces degradation of p27^{Kip1} in pulse-chase labelled fibroblasts. [35S]p27^{Kip1} levels are shown from immunoprecipitations from rMIF-treated (+) versus buffer-treated (-) cells. Right, MIF effect is proteasome-linked. As **a**, but incubations followed by treatment with DMSO, cycloheximide (CHX) or LLnL.

indicate that the mechanism may be enzyme based. In addition, Jab1–p27^{Kip1} complexes might be subject to MIF-mediated conformational modulation. The MIF–Jab1 interaction may provide a molecular basis for several prominent MIF activities. These include the role of MIF as a redox enzyme^{7,19} and regulator of cell differentiation^{20,21}. Notably, thioredoxin/adult T-cell leukaemia-derived factor, another mediator with enzymatic activity, regulates AP-1 activity by redox effects²². The glucocorticoid antagonistic role of MIF⁵ might be connected to its interaction with Jab1. Immunosuppression by glucocorticoids is linked to IκBα transcription²³, and Jab1 and several kinases including IκBα kinase and JNK have been identified in the signalosome complex¹⁷, suggesting that MIF may modulate signalosome-specific kinase activities. We also speculate that inhibition of glucocorticoid action by MIF might occur by modulation of Jab1–steroid receptor interactions²⁴. □

Methods

Yeast two-hybrid screen

The coding region of human MIF was fused in-frame to the GAL4 DNA-binding domain using the pAS2-1 vector (Clontech). With the resulting bait plasmid pMIF–BD, we screened a human fetal brain library (Clontech) by the yeast two-hybrid method using the manufacturer's protocols. Autoactivation by MIF was not detected. We carried out selection on Trp[−] Leu[−] His[−] medium and tested positive clones containing GAL4 DNA activation domain fusion proteins for β-galactosidase activity.

Recombinant proteins, constructs and antibodies

Human rMIF was purified from *Escherichia coli* as described^{13,19}. For [35S]-radiolabelling, rMIF was expressed in *E. coli* in standard minimal medium supplemented with PROMIX (Amersham) and purified as for rMIF¹³. FLUOS–MIF was prepared from biologically active rMIF¹³ by a standard kit (Roche Diagnostics). Construction of the pBK antisense murine MIF plasmid has been described²⁵. The complete coding sequence of Jab1 was obtained by RT-PCR from Jurkat cells. The Jab1 cDNA was cloned into the pCI-neo vector (Promega) for *in vitro* translation and transfection experiments. We generated antisense Jab1 vector by cloning reverse-orientated Jab1 cDNA into this vector. TNF was from R&D Systems. For EGFP fusion proteins, EGFP was fused to the C-terminal of MIF using the pN3-EGFP vector (Clontech); the use of p33–EGFP as a mitochondrial marker²⁶ has been described; and PKCμ(KD)-EGFP shows cytosolic localization (F.-J.J., unpublished data). Polyclonal anti-rMIF antibody was obtained as described⁴, anti-MIF monoclonal antibody was from R&D Systems. Polyclonal anti-phospho-c-Jun (specificity for Ser 63) antibody was from New England Biolabs. Anti-TNF antibody was from R&D, anti-Flag antibody from Kodak, and Cy-2- and Alexa-546-conjugated secondary antibodies from Molecular Probes. All other antibodies were from Santa Cruz Biotechnology.

In vitro transcription/translation reactions

We carried out coupled *in vitro* transcription/translation reactions with TNT T7 reticulocyte lysates in combination with the Transcend translation detection system (Promega). Jab1 and c-Jun were expressed using pCI-neo-Jab1 and pBAT-c-Jun. Expression efficiency of biotin-labelled Jab1 and c-Jun was confirmed by immunoblotting using target protein-specific antibody or streptavidin/horse radish peroxidase conjugate (S-HRP).

Protein–protein interaction

For binding of soluble rMIF to *in vitro* synthesized Jab1, we depleted TNT lysates of endogenous MIF (estimated at 10 nM) by incubation of lysates in ELISA plates coated with anti-MIF antibodies (100 μg per well) for 1 h at 4 °C. Lysates were supplemented with 1 μM rMIF, control buffer, or bovine serum albumin (BSA), and biotin-labelled Jab1 or c-Jun. Reactions were diluted sixfold in phosphate-buffered saline (PBS) containing proteinase inhibitor cocktail (PI; Roche), and rotated at 4 °C for 3 h. Complexes were immunoprecipitated by antibody/protein-A-Sepharose (Amersham-Pharmacia) (1 h each, 4 °C), washed in PBS, and then in 50 mM Tris–HCl, pH 8.0, 170 mM NaCl, 0.5% NP-40 and 50 mM NaF. Coimmunoprecipitated biotinylated proteins were detected by immunoblotting and S-HRP staining. For coprecipitation of biotinylated MIF–EGFP with Jab1, we used streptavidin-conjugated magnetic beads (Dyna) and detected Jab1 with anti-Jab1 antibody after immunoblotting. We carried out immobilization of rMIF on Affigel 10 (Bio-Rad) using the manufacturer's protocol. Binding of Jab1 to beads was analysed in lysates from 293T cells that had been transfected with pCI-neo-Jab1 for 40 h. Cell lysis was done in 25 mM HEPES, pH 7.7, 0.4 M NaCl, 1.5 mM MgCl₂, 2 mM EDTA, 0.5% Triton X100, 3 mM dithiothreitol and PI. Binding and washes were performed in the same buffer except that the NaCl concentration was diluted fourfold. Interaction of endogenous MIF with endogenous Jab1 was analysed in lysates of untreated 293T cells. MIF–Jab1 complexes were precipitated with anti-Jab1 in comparison with control antibody or protein-A-Sepharose alone and immunoblots analysed with anti-MIF antibody.

Tissue culture and transfections

Cell lines were cultured according to standard procedures. Transfections were performed using Superfect (Qiagen). For incubations with rMIF, MIF was added 2 h after addition

of plasmids. We carried out transfections of serum-starved NIH3T3 cells essentially as described¹⁰. Raw cells expressing reduced amounts of endogenous MIF protein were obtained by transfecting cells with the pBK/antisense MIF expression vector²⁵. Control cells were transfected with an empty pBK plasmid. Stably transfected clones were isolated; we used one clone, ΔAS 2.23, expressing less than 50% of MIF content of control cells, in this study.

Uptake studies and fluorescence microscopy

Raw or HeLa cells were incubated with [35S]rMIF, FLUOS-MIF, biotin-MIF (1 μM each), or labelling reagents alone for 1 h, cells washed in ice-cold PBS and 50 mM glycine, 150 mM NaCl, pH 3, and prepared for liquid scintillation counting, fluorescent microscopy, or confocal microscopy by standard procedures. For [35S]rMIF experiments, cells were subjected to subcellular fractionation by differential centrifugation and cytosolic fractions analysed by PD10 gel filtration. For colocalization studies, HeLa cells were transfected (1 μg pMIF–EGFP) and cells incubated for 24 h. Cells were washed, fixed, permeabilized with 0.05% Tween-20, and blocked with 5% goat serum. For Jab1 staining, anti-Jab1 antibody and Alexa-546 staining was used. Samples were analysed with a confocal laser scanning microscope (Leica) using filters for Alexa-546, GFP, and FLUOS emissions.

Activity assays

AP-1 reporter gene activity was measured as described²⁷. pCI-neo-Jab1 or empty plasmid (0.15 μg) and the Tk-LUC-5x12-*o*-tetradecanoyl phorbol acetate (TRE)²⁸, R15-RSV-LacZ reporter constructs and the pEGFP plasmid (0.05 μg of each) were used. Incubations with rMIF were carried out for 18 h. NFκB reporter gene activity measurements in 293T cells were done as described²⁷, with rMIF added for 40 h. The NFκB assay in Raw cells was done as described²⁹. JNK assays were done as described³⁰, with rMIF added for 48 h. We carried out p27^{Kip1} induction experiments and proliferation studies using a published procedure¹⁰. For pulse-chase labelling of p27^{Kip1}, we adapted a described method¹⁰, with synchronized fibroblasts plated at 7 × 10⁵ cells per well, and 1 μM rMIF added at plating, at the pulse and at the chase period.

Received 8 June; accepted 30 August 2000.

- Bernhagen, J., Calandra, T. & Bucala, R. Regulation of the immune response by macrophage migration inhibitory factor: biological and structural features. *J. Mol. Med.* **76**, 151–161 (1998).
- Bozza, M. *et al.* Targeted disruption of migration inhibitory factor gene reveals its critical role in sepsis. *J. Exp. Med.* **189**, 341–346 (1999).
- Calandra, T. *et al.* Protection from septic shock by neutralization of macrophage migration inhibitory factor (MIF). *Nature Med.* **6**, 164–169 (2000).
- Bernhagen, J. *et al.* MIF is a pituitary-derived cytokine that potentiates lethal endotoxaemia. *Nature* **365**, 756–759 (1993).
- Calandra, T. *et al.* MIF as a glucocorticoid-induced modulator of cytokine production. *Nature* **377**, 68–71 (1995).
- Rosengren, E. *et al.* The immunoregulatory mediator macrophage migration inhibitory factor (MIF) catalyzes a tautomerization reaction. *Mol. Med.* **2**, 143–149 (1996).
- Kleemann, R. *et al.* Disulfide analysis reveals a role for macrophage migration inhibitory factor (MIF) as a thiol-protein oxidoreductase. *J. Mol. Biol.* **280**, 85–102 (1998).
- Claret, F.-X., Hibi, M., Dhut, S., Toda, T. & Karin, M. A new group of conserved coactivators that increase the specificity of AP-1 transcription factors. *Nature* **383**, 453–457 (1996).
- Bianchi, E. *et al.* Integrin LFA-1 interacts with the transcriptional co-activator JAB1 to modulate AP-1 activity. *Nature* **404**, 617–621 (2000).
- Tomoda, K., Kubota, Y. & Kato, J.-Y. Degradation of the cyclin-dependent-kinase inhibitor p27^{Kip1} is instigated by Jab1. *Nature* **398**, 160–164 (1999).
- Fields, S. & Song, O. A novel genetic system to detect protein–protein interactions. *Nature* **340**, 245–246 (1989).
- Bacher, M. *et al.* MIF expression in the rat brain—Implications for neuronal function. *Mol. Med.* **4**, 217–230 (1998).
- Bernhagen, J. *et al.* Purification, bioactivity, and secondary structure analysis of mouse and human macrophage migration inhibitory factor (MIF). *Biochemistry* **33**, 14144–14155 (1994).
- Hofmann, K. & Bucher, P. The PCI domain: a common theme in three multiprotein complexes. *Trends Biochem. Sci.* **23**, 204–205 (1998).
- Asano, K. *et al.* Structure of cDNAs encoding human eukaryotic initiation factor 3 subunits. *J. Biol. Chem.* **272**, 27042–27052 (1997).
- Sun, H., Bernhagen, J., Bucala, R. & Lolis, E. Crystal structure at 2.6 Å resolution of human macrophage migration inhibitory factor. *Proc. Natl Acad. Sci. USA* **93**, 5191–5196 (1996).
- Seeger, M. *et al.* A novel protein complex involved in signal transduction possessing similarities to 26S proteasome subunits. *FASEB J.* **12**, 469–478 (1998).
- Karin, M. The regulation of AP-1 activity by mitogen-activated protein kinases. *J. Biol. Chem.* **270**, 166483–166486 (1995).
- Kleemann, R., Kapurniotu, A., Mischke, R., Held, J. & Bernhagen, J. Characterization of catalytic center mutants of macrophage migration inhibitory factor (MIF) and comparison with C81S MIF. *Eur. J. Biochem.* **261**, 753–766 (1999).
- Janahan, A., Williams, J. B., Sanders, L. K. & Nathans, D. Growth factor-induced delayed early response genes. *Mol. Cell. Biol.* **12**, 3919–3929 (1992).
- Wistow, G. J., Shaughnessy, M. P., Lee, D. C., Hodin, J. & Zelenka, P. S. A macrophage migration inhibitory factor is expressed in the differentiating cells of the eye lens. *Proc. Natl Acad. Sci. USA* **90**, 1271–1280 (1993).
- Hirota, K. *et al.* AP-1 transcriptional activity is regulated by a direct association between thioredoxin and Ref-1. *Proc. Natl Acad. Sci. USA* **94**, 3633–3638 (1997).
- Scheinman, R. L., Cogswell, P. C., Lofquist, A. K. & Baldwin, A. S. J. Role of transcriptional activation of IκBα in mediation of immunosuppression by glucocorticoids. *Science* **270**, 283–289 (1995).
- Chauchereau, A., Georgiaki, M., Perrin-Wolff, M., Milgrom, E. & Loosfelt, H. JAB1 interacts with both the progesterone receptor and SRC-1. *J. Biol. Chem.* **275**, 8540–8548 (2000).

25. Waeber, G. *et al.* Insulin secretion is regulated by the glucose-dependent production of islet beta cell macrophage migration inhibitory factor. *Proc. Natl Acad. Sci. USA* **94**, 4782–4787 (1997).
26. Dedio, J., Jähnen-Dechent, W., Bachmann, M. & Müller-Esterl, W. The multiligand-binding protein gC1qR, putative C1q receptor, is a mitochondrial protein. *J. Immunol.* **160**, 3534–3542 (1998).
27. Johannes, F. J. *et al.* Protein kinase C α downregulation of tumor-necrosis-factor-induced apoptosis correlates with enhanced expression of nuclear-factor- κ B-dependent protective genes. *Eur. J. Biochem.* **257**, 47–54 (1998).
28. Angel, P. *et al.* Phorbol ester-inducible genes contain a common cis element recognized by a TPA-modulated transactivating factor. *Cell* **49**, 729–736 (1987).
29. Roger, T. *et al.* Increased AP-1 and NF- κ B activity and increased stability of IL-8 mRNA are implicated in superinduction of IL-8 mRNA by epithelial H292 cells. *Biochem. J.* **330**, 429–435 (1998).
30. Berberich, I. *et al.* Cross-linking CD40 on B cells preferentially induces stress-activated protein kinases rather than mitogen-activated kinases. *EMBO J.* **15**, 92–101 (1996).

Supplementary information is available on Nature's World-Wide Web site (<http://www.nature.com>) or as paper copy from the London editorial office of Nature.

Acknowledgements

We thank S. Schlenker and E. Wagner for experimental assistance; P. Angel for the pBAT-c-Jun and pRSV-c-Jun constructs; and R. Bucala for communicating unpublished data on MIF/Erk induction. We thank A. Billich/Novartis Research Center, Vienna, for testing the effect of anti-MIF antibodies on fibroblast growth. We are grateful to K. Pfizenmaier, F. Vitthum, and K. Sohn for helpful discussions and critical reading of the manuscript. This work was supported by a grant from the Fraunhofer IGB.

Correspondence and requests for materials should be addressed to J.B. (e-mail: jbe@igb.fhg.de).

Cell-cycle-regulated DNA double-strand breaks in somatic hypermutation of immunoglobulin genes

F. Nina Papavasiliou & David G. Schatz

Section of Immunobiology, Howard Hughes Medical Institute, Yale University School of Medicine, Box 208011, 310 Cedar Street, New Haven, Connecticut 06520-8011, USA

Targeted hypermutation of immunoglobulin variable region genes occurs in B cells during an immune response¹, and gives rise to families of related mutant antibodies which are then selected for their binding affinity to the immunizing antigen². Somatic hypermutation predominantly generates point mutations, many of which occur at specific residues (hotspots)³. The reaction has been linked to transcription and requires the presence of immunoglobulin enhancers^{4–6}, but replacement of the variable gene by heterologous sequences, or the variable region promoter by a heterologous promoter, does not interfere with the mutation process^{7,8}. Here we show the existence of abundant DNA double-strand breaks (DSBs) in hypermutating sequences. Generation of the DSBs is coupled to transcription, enhancer-dependent, and correlates with the appearance of nearby mutations. Furthermore, the DSBs are cell-cycle restricted, being found almost exclusively in cells that have completed, or nearly completed, DNA replication. We propose a model for somatic hypermutation in which mutations are introduced into the DNA during repair of DSBs by homologous recombination. The finding of DSBs during somatic hypermutation may help to explain the chromosomal translocations found in some B-cell tumours.

Somatic hypermutation was originally hypothesized to be a two-step process, initiated by cleavage of the DNA within the mutating region, and subsequently resolved by error-prone repair⁹. Direct evidence for the involvement of DNA strand breaks in somatic hypermutation has been elusive¹⁰. Indirect support for the existence

of such breaks comes from experiments describing the constitutively hypermutating Ramos cell line, and the ability of terminal deoxynucleotidyl transferase (TdT)-transfected Ramos cells to accumulate untemplated nucleotide additions near mutational hotspots within immunoglobulin variable regions¹¹.

To investigate the nature of these strand lesions, we probed Ramos genomic DNA for locus-specific DNA DSBs by ligation-mediated polymerase chain reaction (LM-PCR)¹². Blunt DSBs were readily detectable over the Ramos immunoglobulin heavy- and light-chain variable regions (V_H and V_λ) (Fig. 1A, lanes 1 and 2), but could not be amplified from the heavy-chain constant region (lane 11), or various other genomic loci (the GAPDH, B29 and RAG1 genes) (lanes 5, 7, 9). Breaks were not detectable over the V_H region of a mouse myeloma line (J558) (lane 3), which does not hypermutate¹³. The inability to detect DSBs over these non-mutating regions was not due to PCR failure, as DSBs could be amplified from DNA samples spiked with plasmids linearized in the region of interest (lanes 4, 6, 8, 10, 12). With the exception of RAG1, all the genes tested are transcribed in Ramos cells (data not shown), but only the V_H and V_λ regions have been reported to mutate.

To determine the frequency of DSBs over the Ramos immunoglobulin variable regions, we carried out a serial dilution experiment in which linker-ligated DNA was diluted into unligated Ramos DNA (Fig. 1B), keeping the total amount of DNA per amplification reaction constant (Fig. 1Be). Using gene-specific nested primers from the 5' end of the rearranged V_H region, DSBs were detected reproducibly in as few as 50 cells (Fig. 1Ba). Two distinct bands (and hence at least two independent DSBs) were detected from 50 cells, indicating that perhaps 4% of Ramos cells in bulk culture contain DSBs over the rearranged V_H at any given point in time (Fig. 1Ba, lane 8). Furthermore, DSBs could be detected using nested primers from the 5' end of the rearranged V_λ gene, in which case at least one DSB could be amplified from as few as 10 cells (Fig. 1Bc, lane 9). Because Ramos cells carry two rearranged λ light-chain alleles (both of which are detected by our LM-PCR assay), these data indicate that about 5% of Ramos rearranged V_λ alleles may contain a DSB.

We noted that the two ends created by the DSB were not detected with equal frequency. 'Upstream' ends (which we define as those lying on the fragment of DNA containing the promoter, and hence detected with forward orientation primers) of V_H and V_λ DSBs were detected with high frequency (Fig. 1Ba and c). In contrast, 'downstream' ends (defined as those lying on the piece of DNA separated from the promoter, detected with reverse orientation primers) were detected with at least 20-fold lower frequency (Fig. 1Bb and d). Similar results were obtained for the artificial mutation loci described below (data not shown). Treatment of agarose-embedded DNA with T4 DNA polymerase (to blunt ends with 5' or 3' overhangs) or mung bean nuclease (to open covalently sealed, hairpin ends) before linker ligation did not enhance the frequency of detection of downstream ends (data not shown). We conclude that DSB formation in somatic hypermutation results in asymmetric products: upstream ends that are usually blunt, and downstream ends that often contain some modification that prevents their detection. The nature of this modification is currently unknown.

To ascertain the distribution of DSBs and mutations over the Ramos variable regions, we cloned and sequenced LM-PCR and standard PCR products from these regions (the latter from genomic DNA as well as from complementary DNA). For the identification of breakpoints, linker was ligated to six separate DNA samples, followed by LM-PCR and cloning of V_H and V_λ breaks. We sequenced cohorts of 40 clones from each of the 6 amplification reactions, and discarded identical sequences from the same cohort. For both V_H and V_λ DSBs were often intimately associated with mutations, occurring either directly at, or 1–2 nucleotides away from, a mutation (Fig. 1C, D). The DSBs were significantly more likely to lie in the consensus hotspot sequence R-G-Y-W (and its

inverse complement, W-R-C-Y) than in other sequences (see Supplementary Information). A simple interpretation of these data is that a DSB often gives rise to a nearby mutation. Given the frequency and location of the DSBs over the mutating variable regions, we conclude that in Ramos cells, DSBs over the variable regions are tightly linked to somatic hypermutation.

We next wanted to determine whether the appearance of DSBs over mutating loci was dependent on transcription and the presence of immunoglobulin enhancer elements. We therefore engineered an artificial mutation substrate (named MT) that contained the Igk enhancers and immediate-early cytomegalovirus (CMV) promoter driving expression of a neomycin phosphotransferase/green fluorescence protein fusion cassette (neo-GFP), which was modified such that a stop codon (and mutational hotspot) replaced the third codon of the neo gene (Fig. 2A). Mutation of the hotspot could revert the stop codon, allowing expression of the neo-GFP fusion protein and fluorescence-based detection of mutationally active cells. Between the two enhancers, the MT substrate contained a yellow fluorescence protein/puromycin resistance fusion gene (YFP-puro), also under the control of a CMV promoter. On the basis of results with a transgene containing a duplication of the Igk promoter upstream of Ck⁴, the YFP-puro gene would also be expected to hypermutate.

We stably transfected Ramos cells with the MT mutation substrate and with a similar substrate that lacked both enhancers (MTΔE; Fig. 2A). As the neo-GFP and the YFP-puro expression cassettes are under the transcriptional control of the enhancer-independent, constitutive CMV promoter, we could easily detect their transcripts in both transfected Ramos lines (Fig. 2B). However, DSBs over the neo-GFP and YFP-puro genes were detected only if the enhancers were present (Fig. 2C). Therefore, immunoglobulin enhancers are not simply required to drive transcription of the mutating locus, but

rather, they have a central role in initiating the process of hypermutation in the context of transcription.

To test the role of transcription in the generation of DSBs, we modified the MT construct by replacing the CMV sequences upstream of neo-GFP with a tetracycline inducible promoter (tMT; Fig. 2A). Ramos cells were co-transfected with tMT and a vector directing inducible expression of the tetracycline transactivator, tTA¹⁴. When transfected cells were cultured in the presence of tetracycline, neither neo-GFP transcripts (Fig. 2Dc, day 0) nor DSBs over the neo-GFP gene (Fig. 2Dd, day 0) were detectable. In contrast, transcripts (Fig. 2Db) and DSBs (Fig. 2De) were easily detected from the YFP-puro gene under the control of the constitutive CMV promoter. Induction of transcription by removal of tetracycline from the culture medium (Fig. 2Dc) resulted in the accumulation of DSBs over the neo-GFP cassette (Fig. 2Dd). Cloning and sequencing of LM-PCR and PCR products from the neo-GFP fusion gene revealed that mutations were first detected at day 5, one day after the first detection of DSBs, and that at both day 5 and day 6, the majority of mutations were adjacent to breakpoints (data not shown). We conclude that the process of transcription is crucial to the generation of the DSB intermediates, and that in the absence of transcription, the enhancers are not sufficient to drive the production of DSBs over the neo-GFP gene, consistent with the need for a promoter for hypermutation *in vivo*^{4,6}. Mechanistically, the need for transcription can be interpreted as a requirement for (1) the assembly of the transcription initiation complex on the promoter; (2) transcription elongation; or (3) the nascent transcript. Our data do not allow us to distinguish between these possibilities.

It was important to show that our observation of DSBs in the *in vitro* model system could be extended to the *in vivo* situation, that is, splenic B cells after immunization. We therefore created transgenic mice carrying the MT construct and analysed them 14 days after

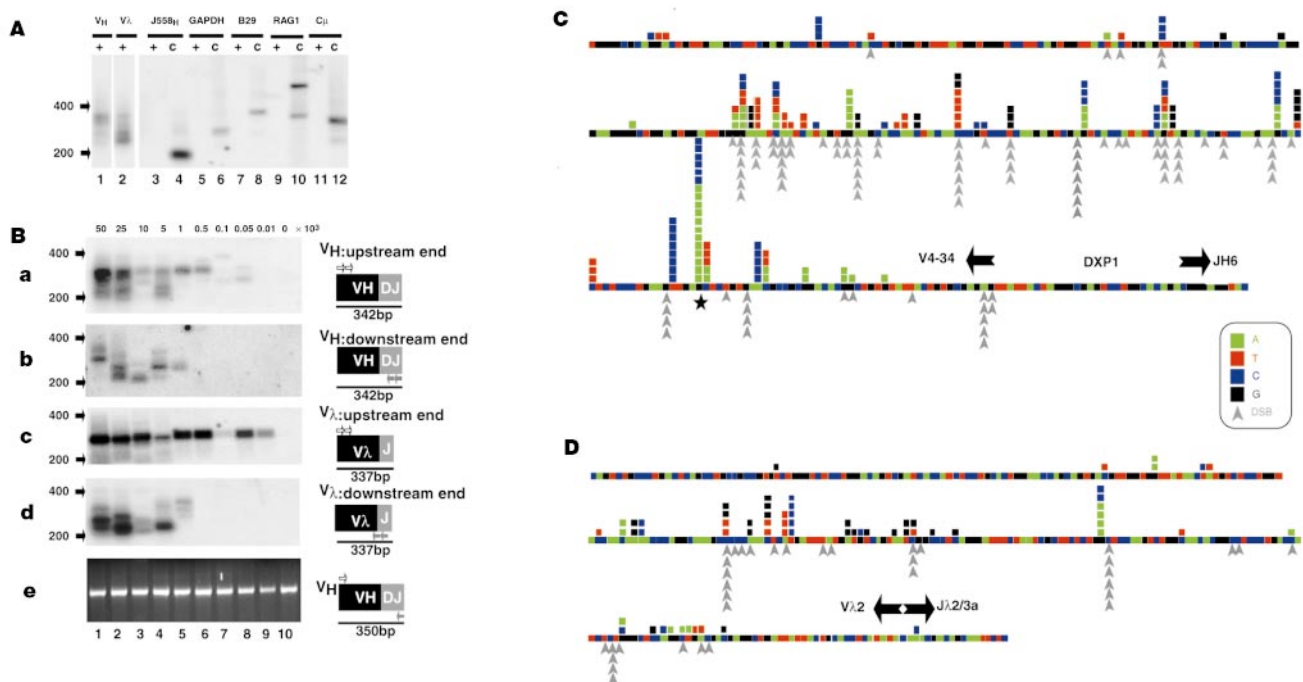


Figure 1 DSBs are tightly linked to mutations over the Ramos immunoglobulin V_H and V_λ regions. **A**, LM-PCR amplification of DSBs from the Ramos genes indicated. Cμ, immunoglobulin heavy chain constant region; J558_H, the V_H region of the J558 myeloma cell line; +, linker-ligated genomic DNA template; c, control linker-ligated template consisting of genomic DNA plus plasmid linearized in the region of interest. **B**, Dilution analysis for quantification of variable region DSBs using the indicated cell equivalents of linker-ligated DNA. **a–d**, LM-PCR analysis. **e**, PCR amplification of the Ramos V_H region.

Primers are indicated as arrows. **C, D**, Location of DSBs and mutations in the rearranged immunoglobulin V_H and V_λ regions. Above each variable region (depicted as a linear series of coloured squares) the nature and position of mutations are indicated; below, arrowheads indicate positions of DSBs by pointing to the variable region nucleotide attached to the LM-PCR linker. Asterisk, hotspot mutated in virtually all sequences obtained. These mutations often convert the sequence into one that should no longer function as a hotspot.

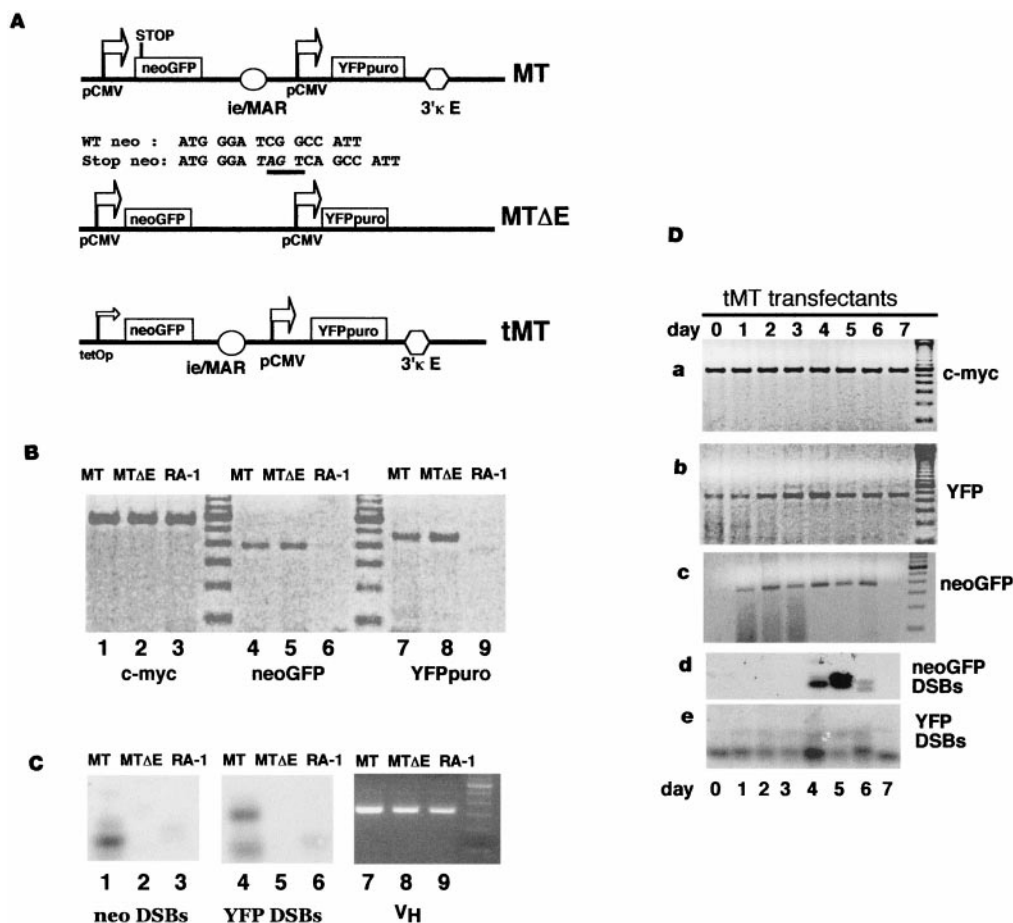


Figure 2 Double-strand breaks are enhancer dependent and require transcription. **A**, Mutation constructs MT, MTΔE and tMT. pCMV, cytomegalovirus immediate-early promoter; ie/MAR, Igκ intronic enhancer and matrix attachment regions; 3'κ E, Igκ 3' enhancer; tetOp, tetracycline-inducible promoter. Initial codons of the neo-GFP (stop neo) and wild-type neo (WT neo) genes are indicated (stop codon in italics; AGT hotspot sequence underlined). **B**, RT-PCR assay of c-Myc, neo-GFP and YFP-puro transcripts in

Ramos transfectants (MT and MTΔE) and untransfected controls (RA-1). **C**, LM-PCR analysis of DSBs (lanes 1–6) and direct PCR amplification of the endogenous V_H gene (lanes 7–9). **D**, RT-PCR detection of transcripts (**a–c**), and LM-PCR amplification of DSBs (**d, e**) in tMT-transfected Ramos cells before induction (day 0) and up to 7 days after induction.

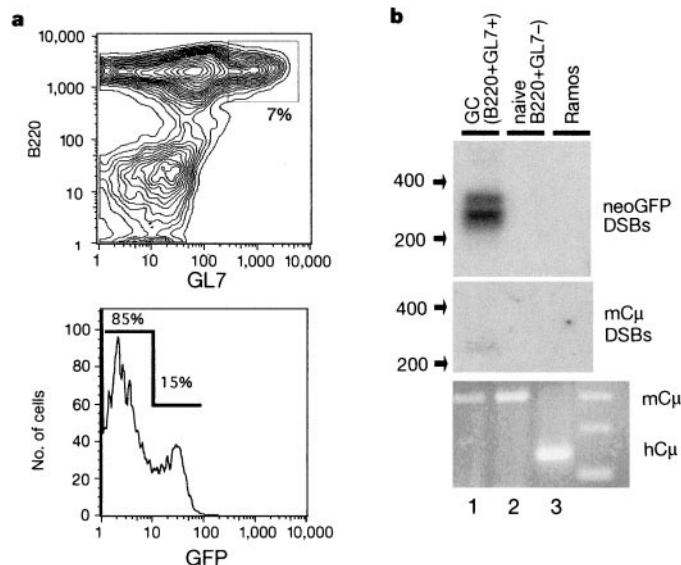


Figure 3 Detection of DSBs in mutating splenic B cells from immunized MT-transgenic mice. **a**, Flow cytometric analysis showing germinal centre B220⁺GL7⁺ cells (contour plot, upper panel) and analysis of this population for intrinsic GFP fluorescence (histogram, lower panel). **b**, Top and middle, LM-PCR analysis of DSBs in sorted splenic B cells (mCμ; mouse heavy chain constant region). Bottom, direct PCR amplification of mCμ or human Cμ (hCμ).

lower panel). **b**, Top and middle, LM-PCR analysis of DSBs in sorted splenic B cells (mCμ; mouse heavy chain constant region). Bottom, direct PCR amplification of mCμ or human Cμ (hCμ).

immunization. Splenic lymphocytes were stained with antibodies specific for B220 (a marker of B cells) and GL7 (a marker of germinal-centre B cells), and B220⁺GL7⁺ B cells were sorted on the basis of GFP fluorescence (Fig. 3a). In different experiments, 2–20% of the B220⁺GL7⁺ germinal-centre cells were GFP⁺ after immunization (Fig. 3a, bottom panel). DSBs over the neo–GFP transgenic locus were abundant in sorted B220⁺GL7⁺ germinal-centre B cells, but were undetectable in B220⁺GL7[−] naive B cells (Fig. 3b, lanes 1 and 2). DSBs were not a result of nonspecific DNA breakage (such as would occur during apoptosis), as we were not able to amplify any products from the C μ region (Fig. 3b, middle panel). Sequence analysis of the neo–GFP fusion gene revealed that GFP⁺ germinal-centre B cells (but not GFP[−] cells) contained mutations that reverted the stop codon at the beginning of the fusion gene (data not shown). We conclude that in germinal-centre B cells *in vivo*, just as in Ramos cells in culture, DSBs occur over hypermutating regions. In addition, it appears that GFP fluorescence can be used to track hypermutated cells, making the MT-transgenic mice a useful tool for the study of hypermutation *in vivo*.

Our results suggest that somatic hypermutation involves the generation and repair of DSBs. The two major pathways of DSB repair, non-homologous end-joining (NHEJ) and homologous recombination, appear to operate predominantly in distinct phases of the vertebrate cell cycle (G₁/early S and late S/G₂,

respectively)^{15–17}. The cell-cycle distribution of DSBs in somatic hypermutation might therefore shed light on the pathway used to resolve the breaks. We therefore stained MT-transfected Ramos cells with propidium iodide and separated them by fluorescence-activated cell sorting (FACS) according to DNA content and hence cell cycle status. Sorting gave rise to three populations enriched in G₁, S or G₂ phase cells (Fig. 4A). Whereas the G₁ and G₂ populations were well separated, each overlapped to a considerable extent with the S-phase population. For four different loci (V_H, V _{λ} , neo–GFP and YFP–puro), DSBs were detected almost exclusively in the S and G₂ populations (Fig. 4Ba–d). No breaks were detected over the C μ locus, indicating that the staining and sorting procedure did not introduce random DSBs (Fig. 4Be). In addition, analysis of the cell-cycle distribution of DSBs in germinal-centre B cells from immunized MT-transgenic mice revealed that DSBs were at least 100-fold more abundant in the G₂ population than in the G₁ population (Fig. 4C). We conclude that both in Ramos cells and in normal mutating

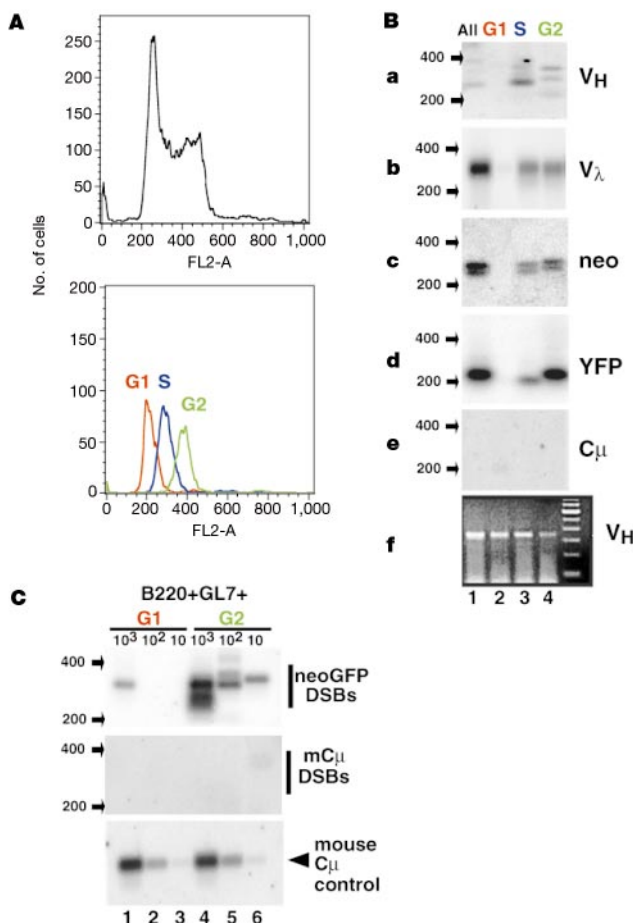


Figure 4 DSBs accumulate predominantly in the late S/G₂ phase of the cell cycle. **A**, Flowcytometric analysis of Ramos cells stained with propidium iodide (top) and re-analysis of the 2n (G₁), intermediate (S) and 4n (G₂) fractions after sorting (bottom). **B**, **a–e**, LM–PCR amplification of DSBs; **f**, direct PCR amplification of the immunoglobulin V_H region. **C**, Top and middle, LM–PCR analysis of DSBs in B220⁺GL7⁺ splenic B cells from immunized MT-transgenic mice (sorted for propidium iodide staining as in **A** above); bottom, direct PCR amplification of the mouse C μ C μ 1 exon.

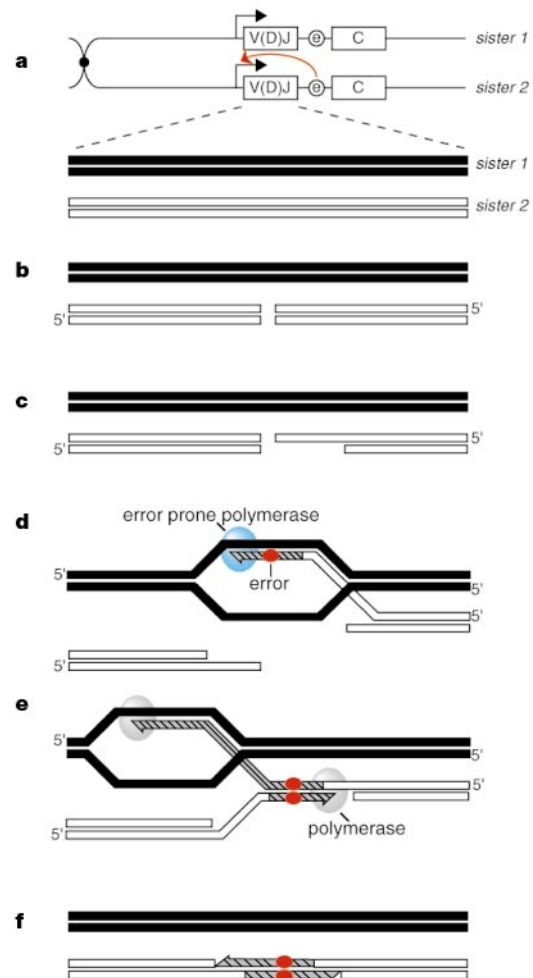


Figure 5 Model for somatic hypermutation. **a**, Transcription (black arrows) and the action (red arrow) of enhancer elements (e) are critical for the initiation of hypermutation. **b**, Introduction of a DSB into one of the two sister chromatids. **c**, 5' end resection. The downstream 5' end is shown resected first, to accommodate the higher efficiency with which we detect upstream ends. **d**, The exposed 3' end invades the sister chromatid and primes new DNA synthesis (hatched lines) by an error prone DNA polymerase (blue oval), which can result in a misincorporation event (red oval). **e**, Strand displacement and second strand synthesis. A more processive, higher fidelity DNA polymerase (gray oval) may be involved. **f**, Ligation. Mutational targeting has been suggested to occur in two stages, one focused on hot spots, and the other, dependent on mismatch repair, yielding a wider distribution of mutations²⁸. Our model does not address the role of mismatch repair in hypermutation. Model derived from refs 29, 30.

cells *in vivo*, DSBs associated with somatic hypermutation are rare or absent in G₁/early S-phase cells and are present in late-S and/or G₂-phase cells.

Our findings strongly implicate DNA DSBs in the process of somatic hypermutation. We propose that a DNA DSB, generated in a transcription-coupled and immunoglobulin enhancer-dependent process, is the initiating event in somatic hypermutation, and hence that the introduction of mutations occurs as a result of the processing of DSBs. The frequency of DSBs in the late S/G₂ phase of the cell cycle leads us to propose that they are usually repaired by homologous recombination, and that this repair process results in mutations through recruitment of an error-prone DNA polymerase (Fig. 5). In the yeast *Saccharomyces cerevisiae*, repair of a DSB by homologous recombination results in a ~100-fold increase in the rate at which point mutations accumulate in the vicinity of the break¹⁸, and these point mutations are almost entirely dependent upon the error-prone polymerase- ζ (ref. 19). Recent studies suggest that pol- ζ acts in concert with pol- ι in DNA lesion bypass, with pol- ι performing the initial extension from the lesion and then being replaced by pol- ζ (refs 20, 21). We propose that a non-processive, low-fidelity polymerase (such as pol- ι) performs the initial extension from hypermutation DSBs (Fig. 5d), resulting in clustering of mutations near the DSB; this polymerase is then replaced (Fig. 5e) by a more processive, higher fidelity polymerase (such as pol- ζ). Either NHEJ or polymerase slippage/mispriming during homologous repair could explain the occasional deletion and insertion events associated with somatic hypermutation^{22,23}.

By linking potentially oncogenic DNA DSBs to the process of somatic hypermutation, our results may explain the genesis of some B-cell malignancies. Many human B-cell lymphomas are of germinal-centre origin, exhibit a high frequency of hypermutation in their immunoglobulin variable genes, and contain chromosomal translocations involving immunoglobulin loci²⁴. On the basis of chromosomal breakpoints, some of these translocations are most easily explained as arising from errors in either V(D)J recombination or switch recombination, two processes known or thought to involve DNA DSBs^{25,26}. Other translocations, however, are not easily accommodated by such models. It has been proposed that these chromosomal translocations arise from DSBs associated with somatic hypermutation²², and our findings provide support for this idea. $\square$

Methods

Mice and cell culture

Transgenic mice were created by injection of the MT construct into C57BL/6 oocytes. Four founder lines were obtained, one of which (MT6282) gave the highest GFP fluorescence intensity in B cells on immunization, and was selected for further study. The human Burkitt lymphoma Ramos (RA-1) cell line was obtained from the ATCC and was grown as described¹¹. All transfected cell lines analysed were polyclonal.

Plasmid constructs

The neo-GFP and the YFP-puro fusion genes (with flanking CMV promoter) were obtained from Clontech. A missense mutation introduced into YFP (Asn121→Ile), renders the YFP-puro protein nonfluorescent. The ie/MAR fragment (*HindIII*-*HpaI* of GenBank sequence V00777) and the 3' κ enhancer (*EcoRI*-*XhoI* fragment, GenBank sequence X15878) were subcloned from a full length Ig κ construct (a gift from M. Shlomchik).

LM-PCR analysis of DNA DSBs

Cells were embedded in agarose and genomic DNA prepared as described²⁷. Linker ligation was done on DNA from 5×10^5 cells in a 50- μ l reaction using the double-stranded anchor linker described previously¹². PCR amplifications were done in two rounds: 10 cycles with a 50°C annealing step (64°C for upstream V λ ends), followed by 28 cycles with a 64°C annealing step (1 μ l of the 50- μ l first round reaction was used to prime the 25- μ l second round reaction). Amplification reactions were done with hotstart Taq polymerase (Qiagen) and sets of linker-specific¹² and nested, gene-specific primers. Reaction products were separated on 2% agarose gels, blotted onto GeneScreen membranes and hybridized to ³²P- γ -ATP labelled gene-specific oligonucleotide probes. Input DNA control reactions amplified either the Ramos V λ or the mouse or human C μ regions in a standard 30-cycle PCR reaction. For the sequences of gene specific primers and the oligo probes, see Supplementary Information.

PCR with reverse transcription (RT-PCR)

Total Ramos RNA was prepared using RNAsol (Tel-Test, Inc.), treated with DNase, and reverse transcribed using oligo-(dT) primers. c-Myc, neo-GFP and YFP-puro sequences were amplified from cDNA from 10^4 cells in a 30-cycle PCR reaction using hotstart Taq polymerase and combinations of appropriate primers (see Supplementary Information).

Flow cytometry and cell sorting

To purify germinal-centre B cells, we pooled spleens from mice at 14 days post-immunization with NP-CGG in alum (NP₂₂:CGG conjugation ratio; from Biosearch Technologies), and separated the activated lymphocytes on a Percoll gradient (Pharmacia). Cells were stained with the pan-B cell marker B220 (phycoerythrin) and the B/T activated cell marker GL7(biotin) (both from BD-Pharmingen), counterstained with streptavidin RED670 (Gibco-BRL), and sorted on a Becton-Dickinson FACS Vantage. Sorted cells were immediately embedded in agarose for further analysis. For the cell-cycle study, cells (Ramos cells or B220/GL7 stained activated B cells) were fixed in 70% ethanol for 30 min at -20°C, and then incubated for 30 min at 37°C in a solution containing 100 μ g ml⁻¹ propidium iodide (Molecular Probes), 20 μ g ml⁻¹ RNaseA and 1 μ l ml⁻¹ NP40 detergent. All analyses of flowcytometric data were performed using the Flowjo software package (Trestar).

Received 31 July; accepted 2 October 2000.

- Rajewsky, K. Clonal selection and learning in the antibody system. *Nature* **381**, 751–758 (1996).
- Weigert, M., Cesari, L., Yonkovich, S. & Cohn, M. Variability in the lambda light chain sequences of mouse antibody. *Nature* **228**, 1045–1047 (1970).
- Neuberger, M. S. *et al.* Monitoring and interpreting the intrinsic features of somatic hypermutation. *Immunol. Rev.* **162**, 107–116 (1998).
- Peters, A. & Storb, U. Somatic hypermutation of immunoglobulin genes is linked to transcription initiation. *Immunity* **4**, 57–65 (1996).
- Goyenchea, B. *et al.* Cells strongly expressing Ig κ transgenes show clonal recruitment of hypermutation: a role for both MAR and the enhancers. *EMBO J.* **16**, 3987–3994 (1997).
- Fukita, Y., Jacobs, H. & Rajewsky, K. Somatic hypermutation in the heavy chain locus correlates with transcription. *Immunity* **9**, 105–114 (1998).
- Yelamos, J. *et al.* Targeting of non-Ig sequences in place of the V segment by somatic hypermutation. *Nature* **376**, 225–229 (1995).
- Tumas-Brundage, K. & Manser, T. The transcriptional promoter regulates hypermutation of the antibody heavy chain locus. *J. Exp. Med.* **185**, 239–250 (1997).
- Brenner, S. & Milstein, C. Origin of antibody variation. *Nature* **211**, 242–243 (1966).
- Lo, A. K., Ching, A. K., Lim, P. L. & Chui, Y. L. Strand breaks in immunoglobulin gene hypermutation. *Ann. NY Acad. Sci.* **815**, 432–435 (1997).
- Sale, J. E. & Neuberger, M. S. TdT-accessible breaks are scattered over the immunoglobulin variable domain in a constitutively hypermutating B cell line. *Immunity* **9**, 859–869 (1998).
- Schlissel, M., Constantinescu, A., Morrow, T., Baxter, M. & Peng, A. Double-strand signal sequence breaks in V(D)J recombination are blunt, 5'-phosphorylated, RAG-dependent, and cell cycle regulated. *Genes Dev.* **7**, 2520–2532 (1993).
- Weiss, S. & Wu, G. E. Somatic point mutations in unrearranged immunoglobulin gene segments encoding the variable region of lambda light chains. *EMBO J.* **6**, 927–932 (1987).
- Vignali, D. A., Carson, R. T., Chang, B., Mittler, R. S. & Strominger, J. L. The two membrane proximal domains of CD4 interact with the T cell receptor. *J. Exp. Med.* **183**, 2097–2107 (1996).
- Jeggo, P. A. Studies on mammalian mutants defective in rejoining double-strand breaks in DNA. *Mutat. Res.* **239**, 1–16 (1990).
- Hendrickson, E. A. Insights from model systems: Cell-cycle regulation of mammalian DNA double-strand-break repair. *Am. J. Hum. Genet.* **61**, 795–800 (1997).
- Takata, M. *et al.* Homologous recombination and non-homologous end-joining pathways of DNA double-strand break repair have overlapping roles in the maintenance of chromosomal integrity in vertebrate cells. *EMBO J.* **17**, 5497–5508 (1998).
- Strathern, J. N., Shafer, B. K. & McGill, C. B. DNA synthesis errors associated with double-strand-break repair. *Genetics* **140**, 965–972 (1995).
- Holbeck, S. L. & Strathern, J. N. A role for REV3 in mutagenesis during double-strand break repair in *Saccharomyces cerevisiae*. *Genetics* **147**, 1017–1024 (1997).
- Johnson, R. E., Washington, M. T., Haracska, L., Prakash, S. & Prakash, L. Eukaryotic polymerases ι and ζ act sequentially to bypass DNA lesions. *Nature* **406**, 1015–1019 (2000).
- Tissier, A. *et al.* Misinsertion and bypass of thymine-thymine dimers by human DNA polymerase ι . *EMBO J.* **19**, 5259–5266 (2000).
- Goossens, T., Klein, U. & Kuppers, R. Frequent occurrence of deletions and duplications during somatic hypermutation: implications for oncogene translocations and heavy chain disease. *Proc. Natl Acad. Sci. USA* **95**, 2463–2468 (1998).
- Wilson, P. C. *et al.* Somatic hypermutation introduces insertions and deletions into immunoglobulin V genes. *J. Exp. Med.* **187**, 59–70 (1998).
- Klein, U. *et al.* Somatic hypermutation in normal and transformed human B cells. *Immunol. Rev.* **162**, 261–280 (1998).
- Kenter, A. L. The liaison of isotype class switch and mismatch repair: an illegitimate affair. *J. Exp. Med.* **190**, 307–310 (1999).
- Fugmann, S. D., Lee, A. I., Shockett, P. E., Villy, I. J. & Schatz, D. G. The RAG proteins and V(D)J recombination: complexes, ends, and transposition. *Annu. Rev. Immunol.* **18**, 495–527 (2000).
- Petrie, H. T., Livak, F., Burtrum, D. & Mazel, S. T cell receptor gene recombination patterns and mechanisms: cell death, rescue, and T cell production. *J. Exp. Med.* **182**, 121–127 (1995).
- Rada, C., Ehrenstein, M. R., Neuberger, M. S. & Milstein, C. Hot spot focusing of somatic hypermutation in MSH2-deficient mice suggests two stages of mutational targeting. *Immunity* **9**, 135–141 (1998).
- McGill, C. B., Holbeck, S. L. & Strathern, J. N. The chromosome bias of misincorporations during double-strand break repair is not altered in mismatch repair-defective strains of *Saccharomyces cerevisiae*. *Genetics* **148**, 1525–1533 (1998).
- Paques, F. & Haber, J. E. Multiple pathways of recombination induced by double-strand breaks in *Saccharomyces cerevisiae*. *Microbiol. Mol. Biol. Rev.* **63**, 349–404 (1999).

Supplementary information is available on Nature's World-Wide Web site (<http://www.nature.com>) or as paper copy from the London editorial office of Nature.

Acknowledgements

We thank T. Taylor and G. Tokmoulina for help with cell sorting; E. Hilton for special assistance with DNA sequencing; C. Arthur for her role in the creation of transgenic mice; and S. Fugmann and M. Diaz for many helpful discussions. We are very grateful to the following people for helpful comments on the manuscript: M. Diaz, S. Fugmann, J. Haber, D. Hesslein, M. Jasin, M. Nussenzweig and I. Villey. Oligonucleotide synthesis and DNA sequencing were performed by the W. M. Keck Foundation Biotechnology Resource Laboratory at Yale University. F.N.P. was supported by a postdoctoral fellowship from the Arthritis Foundation and D.G.S. is an associate investigator of the Howard Hughes Medical Institute.

Correspondence and requests for materials should be addressed to D.G.S. (email: david.schatz@yale.edu).

PCNA connects DNA replication to epigenetic inheritance in yeast

Zhiguo Zhang, Kei-ichi Shibahara* & Bruce Stillman*

Cold Spring Harbor Laboratory, One Bungtown Road, Cold Spring Harbor, New York 11724, USA

Formation of a heterochromatin-like structure results in transcriptional silencing at the *HM* mating-type loci and telomeres in *Saccharomyces cerevisiae*^{1–3}. Once formed, such epigenetically determined structures are inherited for many mitotic divisions⁴. Here we show that mutations in the proliferating cell nuclear antigen (PCNA), an essential component at the DNA replication fork⁵, reduced repression of genes near a telomere and at the silent mating-type locus, *HMR*. The *pol30-8* mutant displayed coexistence of both repressed (pink) and de-repressed (white) cells within a single colony when assayed with the *ADE2* gene inserted at *HMR*. Unlike *pol30-8*, the *pol30-6* and *pol30-79* mutants partially reduced gene silencing at telomeres and the *HMR* and synergistically decreased silencing in cells lacking chromatin assembly factor 1 (CAF-1). All silencing defective mutants showed reduced binding to CAF-1 *in vitro* and altered chromatin association of the CAF-1 large subunit *in vivo*. Thus, PCNA participates in inheritance of both DNA and epigenetic chromatin structures during the S phase of the cell cycle, the latter by at least two mechanisms.

PCNA is important in many aspects of DNA metabolism, including DNA replication and DNA repair in eukaryotes⁵. Human PCNA interacts with CAF-1, a three-subunit protein⁶, to couple DNA replication or DNA repair to nucleosome deposition^{7,8}. In yeast, deletion of any genes encoding CAF-1 subunits (*CAC1*, 2 and 3) results in partial loss of silencing at telomeres and silent mating-type loci^{9,10}. Furthermore, mutations in the *Drosophila* PCNA gene *mus209* suppress position effect variegation¹¹. We therefore tested whether mutations in yeast *S. cerevisiae* PCNA affected silencing.

The *URA3* gene, which is integrated near the telomere of chromosome VII, was used to assay telomere position effect variegation (TPE)¹². Owing to variegated expression of the *URA3* gene at the telomere, wild-type cells grew efficiently in medium lacking uracil, an indication of expression of the *URA3* gene, and medium containing the drug FOA, which kills cells expressing the *URA3* gene (Fig. 1a, *Pol30-WT*). When repression of the *URA3* gene was reduced by deletion of the *CAC1* gene, cells exhibited reduced growth in FOA-containing medium⁹ (Fig. 1a, *cac1Δ*). We tested the

effect of 24 viable, site-specific PCNA mutants (*pol30*) on TPE (see Supplementary Information). Six mutant alleles, *pol30-6*, *pol30-8*, *pol30-79*, *pol30-41*, *pol30-42* and *pol30-45*, reduced TPE significantly (Fig. 1a; Table 1 in Supplementary Information), whereas the remaining 18 mutants did not show detectable effects on TPE compared to the wild type and were not analysed further (for example, *pol30-1* in Fig. 1a; Table 1 in Supplementary Information). We focus here on three alleles, *pol30-6*, *pol30-8* and *pol30-79*. Yeast colony sizes of wild-type *POL30* and these three mutants were monitored when grown in FOA-containing medium. In the absence of FOA, all cells grew similarly, but in FOA, these mutants exhibited reduced colony size compared with the wild type (Fig. 1b). The reduction in the number and size of colonies for these *pol30* mutants indicated that these mutants were defective in TPE.

To study whether *pol30* mutations also affected silencing at *HMR*, the reporter gene *ADE2* inserted at the *HMR* locus was used to detect changes in *HMR* repression by monitoring the colour of the colonies. When the *ADE2* gene was repressed at the *HMR* locus, the colonies were pink. When the *ADE2* gene was completely de-repressed by mutations in genes required for silencing, such as the four silent information regulator (SIR) proteins¹, the colonies were white¹³. All six mutants that affected TPE showed various degrees of reduction in *HMR* repression, whereas three randomly tested mutants that retained TPE had no effect on *HMR* repression (Fig. 2a and Table 1 in Supplementary Information). The colony colour of *pol30-6* and *pol30-42* was close to white and uniform for all of the colonies, indicating that these mutants completely de-repressed the *ADE2* gene. However, these two mutants grew more slowly than the wild-type strain and it was therefore possible that slow growth indirectly affected *HMR* silencing. The colony colour of *pol30-79* was uniform and light pink for all the colonies, and it was close to the colour of *cac1Δ* strains (Fig. 2a). Thus the *pol30-79* allele only partially de-repressed *HMR*.

In contrast to the *pol30-6* and *pol30-79* mutants, pink and white sectorized colonies were observed for three alleles, *pol30-8*, *pol30-41* and *pol30-45*. Within the population, however, 5–10% of colonies were light pink and about 5–20% of colonies were mainly white in each of these three mutants (Table 1 in Supplementary Information). When cells from a white *pol30-8* colony were re-plated

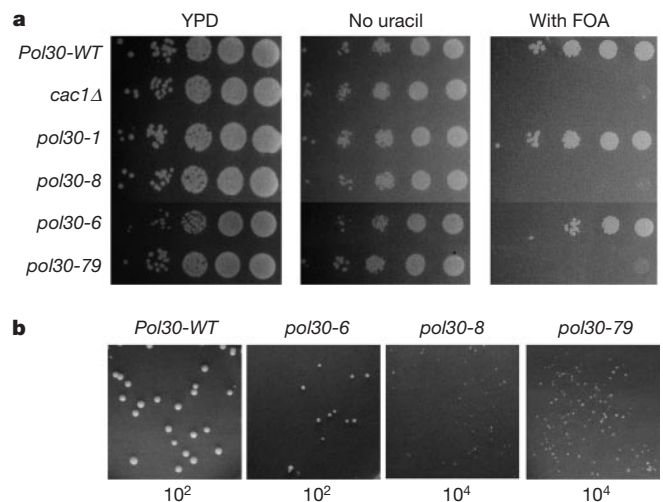


Figure 1 The effect of PCNA mutations on telomeric silencing. **a**, PCNA mutations reduced telomeric silencing. Tenfold serial dilutions of freshly grown cells were plated onto either YPD, SCM-Ura (no uracil) or SCM+FOA (with FOA) medium. Strains shown were *Pol30-WT* (ZGY003-WT), *cac1Δ* (PKY106), *pol30-1* (ZGY003-1), *pol30-8* (ZGY003-8), *pol30-6* (ZGY003-6) and *pol30-79* (ZGY003-79). **b**, Colony size of wild-type *POL30* or *pol30* mutants grown on FOA plates. Freshly grown yeast cells were spread onto FOA plates, and incubated for four days at 30 °C. Number of cells spread indicated below were based on initial dilution. Yeast strains shown were ZGY003-*x* (*x* = WT, 6, 8, 79).

* Present address: TORENT Japan Science and Technology Corporation, Center for Molecular Biology and Genetics, Kyoto University, Sakyo-ku, Kyoto 606-8501, Japan.

immediately onto yeast-peptone-dextrose (YPD) plates, about 50% of progeny colonies remained white. After growth in liquid medium before assaying for *ADE2* expression, the white colony progeny gradually converted to pink/white sectorial colonies (Fig. 2b and c). Similar results were obtained with the predominantly pink *pol30-8* colonies (Fig. 1 in Supplementary Information). These results suggested that the two states, repressed (pink) and de-repressed (white), could be inherited, but were not stable. The coexistence of two transcriptional states in a single cell culture is reminiscent of the phenotype exhibited by deletions of the origin recognition complex (ORC) binding sites at the *HML* or *HMR* silencers or in cells lacking Sir1^{4,13,14}. We conclude that PCNA is required for epigenetic inheritance and establishment of silent states of chromatin.

The effect on TPE and *HMR* repression was also examined when each of the three mutants, *pol30-6*, *pol30-8* and *pol30-79*, was combined with *cac1Δ*. As shown in Table 1, *pol30-8 cac1Δ* did not cause any significant synergistic reduction in TPE or *HMR* repression. One mechanism for PCNA to function in epigenetic inheritance may therefore involve interaction with CAF-1. In contrast, each of the double mutants *pol30-79 cac1Δ* and *pol30-6 cac1Δ* synergistically decreased TPE, but to a different extent (Table 1). The *pol30-79 cac1Δ* also greatly reduced silencing at *HMR* as assayed by colony colour, but we could not determine whether *pol30-6 cac1Δ* decreased *HMR* silencing because the colony colour of *pol30-6* was already white (Table 1). These genetic results suggest that *pol30-79* and *pol30-6* belong to the same group of *pol30* mutants that affect silencing independently of CAF-1. Moreover, our data in Figs 2 and 3 of the Supplementary Information indicated that silencing defects exhibited by the *pol30* mutants may not be secondary effects arising from either DNA repair defects or changes in telomere length.

To determine whether yeast PCNA physically interacts with yeast CAF-1, the carboxy terminus of Cac2 was tagged with the myc epitope (Cac2-13Myc). Precipitation of Cac2-13Myc co-precipitated PCNA, but no PCNA was co-precipitated in cells lacking Cac1 (Fig. 3a, upper panel) even though equal amounts of Cac2 were

precipitated in both cases (Fig. 3a, lower panel). These results suggested that PCNA interacted with CAF-1 through Cac1. To test this hypothesis *in vitro*, GST-Cac1 was immobilized onto glutathione-sepharose beads, and used to bind and co-precipitate (pull down) *in vitro*-translated, ³⁵S-methionine-labelled PCNA (Pol30). GST-Cac1 efficiently pulled down the wild-type PCNA in a concentration-dependent manner (Fig. 3b). Under the same conditions GST-REGα, a proteasome regulator, did not bind PCNA. Thus the interaction between PCNA and Cac1p was conserved from yeast to human^{7,8}.

We also tested whether the PCNA mutants that were defective in silencing affected their association with Cac1 *in vitro* using the GST pull-down assay. When lower amounts of GST-Cac1 were used, none of the four PCNA mutants, Pol30-6, Pol30-8, Pol30-42 and Pol30-79 bound to GST-Cac1, but under the same conditions, wild-type PCNA did interact (Fig. 3c). When twice the amount of GST-Cac1 was used, the Pol30-8 bound weakly to GST-Cac1 compared to the wild type, but the other three PCNA mutants still did not bind (Fig. 3d). Thus, PCNA mutants defective in silencing showed

Table 1 Synergistic effect of PCNA mutations with *cac1Δ* on silencing

<i>Pol30</i> allele	Silencing at telomere (FOA ^a)		Silencing at <i>HMR</i> (colour)	
	<i>CAC1</i>	<i>cac1Δ</i>	<i>CAC1</i>	<i>cac1Δ</i>
WT	0.65 [0.49–0.92]	2.60×10^{-3} [($1.1-5.3$) $\times 10^{-3}$]	P	LP
6	0.12 [0.04–0.28]	2.5×10^{-5} [($0.79-6.2$) $\times 10^{-5}$]	W	W
8	7.0×10^{-3} [($1.8-12$) $\times 10^{-3}$]	2.9×10^{-3} [($1.4-6.9$) $\times 10^{-3}$]	S	S
79	5.3×10^{-2} [($1.3-8$) $\times 10^{-2}$]	$\leq 1.5 \times 10^{-6}$ [$\leq (0.81-2.2) \times 10^{-6}$]	LP	W

The *POL30* or mutants *pol30-6*, *pol30-8* and *pol30-79*, or combinations of *cac1Δ* with each of the mutants were tested for their effect on telomere silencing by determining the fraction of cells resistant to FOA (columns 2 and 3), and on *HMR* repression by colony colour (columns 4 and 5). The fractions of cells resistant to FOA were reported as the median and range of values (in square brackets) from at least six independent experiments. Colony colour was classified into P (pink), LP (light pink), W (white); S (LP/W sectorial). Strains used were ZGY003-x, ZGY007-x, ZGY005-x, and ZGY008-x, where x = WT, 6, 8 or 79.

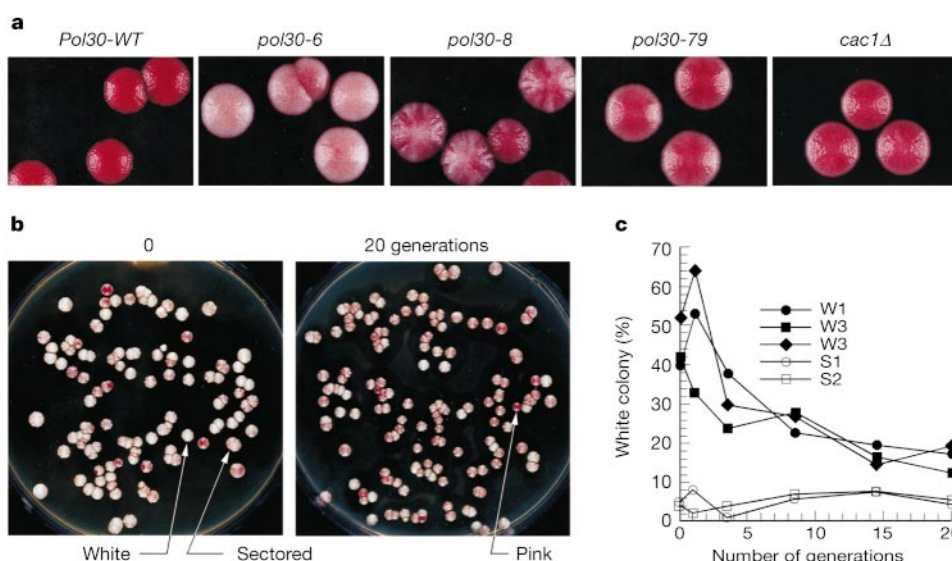


Figure 2 PCNA mutants affected *HMR* repression. **a**, A qualitative colony colour assay was used to assay repression of the *ADE2* gene inserted at *HMR*. Freshly grown yeast cells were spread onto YPD plates at about 100 cells per plate. The plates were incubated at 30 °C for three days, and photographs were taken after additional three days incubation at 4 °C. Yeast strains shown were *Pol30-WT*, *pol30-6*, *pol30-8* and *pol30-79* that corresponded to ZGY005-x (x = WT, 6, 8, 79), and *cac1Δ* (ZGY009). **b**, Conversion of 'white' colonies to sectorial colonies of *pol30-8* mutant. A 'white' colony was picked and

spread immediately onto YPD plates (0) or cultured for about 20 generations (20) before assaying the expression of the *ADE2* gene. **c**, Percentage of 'white' progenies decreased with time from 'white' colonies of the *pol30-8* mutant. Three 'white' (W1–W3) and two sectorial (S1–S2) colonies of *pol30-8* were picked to assay expression of the *ADE2* gene with time (x-axis) and percentage of 'white' of the whole population (y-axis) was determined at the indicated time.

reduced binding affinity for Cac1 *in vitro*.

To gain further insight into the interactions between PCNA and CAF-1, we performed a chromatin-binding assay¹⁵. Yeast cells grown to early log phase were harvested for fluorescence-activated cell sorting (FACS) analysis and for the chromatin-binding assay. Most of the PCNA mutants except *pol30-6* and *pol30-42* exhibited almost identical FACS profiles to the wild-type strain (Fig. 4a), suggesting that these mutants did not alter cell cycle progression. Spheroplasts of wild-type or *pol30* mutants were fractionated into soluble or chromatin-containing fractions following a standard protocol, and Orc3 was used as a control for the chromatin-binding assay¹⁵. The majority of Orc3 bound to chromatin, and only very small amounts existed in the soluble fraction (compare lane 1 to lane 10 of Fig. 4e). Most of the wild-type PCNA was in the soluble fraction, but a small portion was in the chromatin fraction (Fig. 4d, compare lane 1 to lane 10). Under the same conditions, about an equal amount of Cac1-3HA was detected on chromatin and in the soluble fraction. Most of the chromatin-bound Cac1-3HA was processed into two smaller fragments (Fig. 4b, lane 1 and lane 10), indicating that Cac1-3HA was more susceptible to proteolysis on chromatin.

PCNA or mutant proteins bound to chromatin as efficiently as wild-type protein (Fig. 4d, lanes 1–9). However, the *pol30* mutants that reduced silencing significantly decreased chromatin association of Cac1-3HA (Fig. 4b). Conversely, the *pol30-1* and *pol30-22* alleles that had no detectable silencing defects did not reduce chromatin binding of Cac1. In contrast to Cac1, most of the Cac2-13Myc was

in the soluble fraction, but a small fraction was chromatin bound (Fig. 4c, lane 1 and lane 10). Moreover, PCNA mutations had little effect on the ability of Cac2-13Myc protein to bind chromatin (Fig. 4c). Cac2 may be part of other chromatin-bound complexes or Cac1 may dissociate from Cac2 during chromatin sample preparation. We note that an excess of hCAF-1 p60 subunit compared with the CAF-1 p150 subunit exists in human cells¹⁶. These results demonstrated that PCNA and CAF-1 bound to chromatin in yeast and that mutations in *POL30* that affected TPE and mating-type silencing (whether they synergize with *cac1Δ* or not) were defective in stable binding of Cac1 to chromatin.

The locations of the altered amino acids in the three silencing-defective PCNA mutants *pol30-6*, *pol30-8* and *pol30-79* are shown on the three-dimensional structure of the PCNA trimer¹⁷ (Fig. 5). Residues Leu 126 and Ile 128 (mutated in *pol30-79*) and residues Asp 41 and Asp 42 (mutated in *pol30-6*) are located on the inter-domain connector loop and the central loop, respectively. These two loops are part of one surface of PCNA that interacts with DNA replication proteins. In fact, biochemical studies have shown that Pol30-6 and Pol30-79 are defective in association with proteins involved in DNA replication such as Polδ^{18,19}. In contrast, the altered amino acids in *pol30-8*, Arg 61 and Asp 63, are located at the tip of a bulge of the surface of the PCNA trimer (Fig. 5), and they are far removed from regions of PCNA known to interact with other proteins, or amino acids at the trimer interfaces. Indeed, Pol30-8 associates with Polδ, Pole and RFC, a protein complex responsible for loading PCNA onto DNA, like wild-type PCNA¹⁸. Thus, the unique location of the altered amino acids in *pol30-8* on the three-dimensional structure of PCNA provides a structural basis for its differential effects on silencing than *pol30-6* and *pol30-79*.

In *S. cerevisiae*, establishment of HM repression from a non-repressed state requires passage through the S phase of the cell

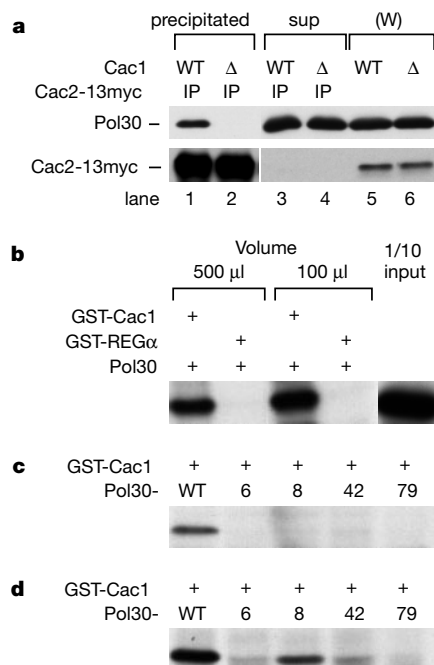


Figure 3 Interactions between PCNA and CAF-1. **a**, Cac2 interacts with PCNA in the presence of Cac1 *in vivo*. Cac2-13Myc was immunoprecipitated from yeast whole-cell extracts prepared from wild-type *CAC1* (WT) or *cac1Δ* (Δ) strains. The precipitated proteins (precipitated) were analysed by antibodies against PCNA (lanes 1 and 2, upper panel), and Cac2-13Myc (lanes 3 and 4, lower panel). Fractions of supernatants (sup) from Cac2-13Myc immunoprecipitation (lanes 3 and 4) and whole cell extracts (W) (lanes 5 and 6) were also analysed. **b**, GST–Cac1 bound to wild-type PCNA *in vitro*. GST–Cac1 or GST–REGα was incubated with *in vitro* translated wild-type PCNA at incubation volumes indicated at the top of **b**, and the bound proteins were resolved by a 12.5% SDS polyacrylamide gel (SDS–PAGE). **c**, GST–Cac1 was incubated with approximately equal amounts of Pol30 and Pol30-6, Pol30-8, Pol30-42 and Pol30-79. All of the bound proteins were analysed by SDS–PAGE. **d**, Double amounts of GST–Cac1 were used for binding assay as described in **c**. + represents the presence of the indicated protein.

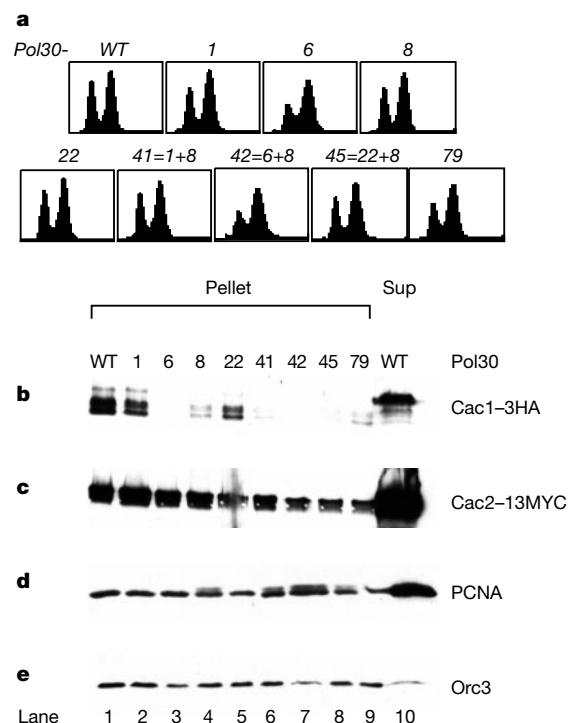


Figure 4 Chromatin association of Cac1 in wild-type PCNA or mutant strains. Yeast cells (ZGY004-x; x = WT, 1, 6, 8, 22, 41, 42, 45, 79) were grown to early logarithmic phase and analysed by FACS (**a**). **b–e**, Proteins from the yeast cells were separated into soluble (Sup) and pellet (chromatin) fractions¹⁵, resolved by 12.5% SDS–PAGE and probed with antibodies against Cac1-3HA, Cac2-13Myc, PCNA and Orc3. We note that the soluble fraction from wild-type PCNA (Sup) was analysed on the same gel for comparison.

cycle²⁰. It was suggested that the interaction between Sir1 and Orc1, a subunit of the ORC, during the S phase was responsible for this process²¹. However, tethering Sir1 to the HMR-E silencer²², a *cis*-acting sequence required for silencing¹, although bypassing the requirements for the ORC binding site and initiation of DNA replication at the silencer, still demands S-phase passage to establish silencing²³. This suggests that the interaction between Sir1 and Orc1 may have other roles in silencing. Our results demonstrate that PCNA, a protein essential in the S phase of the cell cycle⁵, is involved in inheritance of silencing, thereby providing another potential link between DNA replication and epigenetic inheritance. Because combining *pol30-8* and *cac1Δ* did not show a synergistic effect on reduction of silencing at *HMR* or a telomere, we suggest that PCNA and Cac1 function together for establishment of silencing. Consistent with this view, Pol30-8 showed reduced binding to Cac1 *in vitro* and prevented stable binding of Cac1 on chromatin *in vivo*. PCNA interacts with CAF-1 to assemble nucleosomes during DNA replication and repair in mammalian cells⁷, and CAF-1 in turn binds proteins that are components of heterochromatin, such as HP1-like proteins²⁴. We suggest that in yeast, CAF-1 may interact with Sir proteins, and through interactions with PCNA, may help to assemble nucleosomes and also imprint chromatin with an epigenetic mark.

Our results also indicate that PCNA may also be involved in silencing independently of CAF-1, as two PCNA mutant alleles, *pol30-6* and *pol30-79*, were more defective in TPE and *HMR* repression in the absence of CAF-1. It is possible that in addition to CAF-1, PCNA may also load another factor or factors during S-phase to inherit silent chromatin. We suggest that Asf1 is a candidate²⁵. Deletion of *ASF1* and *CAC1* synergistically reduces TPE and *HMR* repression²⁶. The Asf1 orthologue in *Drosophila* forms a complex (called RCAF) with acetylated histones, H3 and H4²⁶, a property shared by CAF-1²⁷. Furthermore, we have found that the *ASF1* interacts genetically with *POL30* in silencing (Z.Z. and B.S., unpublished observation).

A recent report described the restoration of silencing by the *pol30-52* allele and other replication mutants at an *HMR* locus containing a defective silencer²⁸. But in agreement with our data, *pol30-52*, but not other mutants, reduced silencing at an unperturbed *HML* locus²⁸. As Pol30-52 is a monomer, but wild-type PCNA and each of the mutants that we analysed form trimers¹⁸, the *pol30-52* probably affects all aspects of PCNA function *in vivo*. Along with *pol30-52*, other DNA replication mutants that restored silencing at the mutated *HMR* silencer displayed a cold-sensitive growth phenotype²⁸. Thus the improved silencing shown by *pol30-52* may be due to an altered cell cycle progression, allowing more time for cells to establish silencing at the mutant silencer. Mutations in several genes perturbing the cell cycle progression improve *HMR* silencing²⁹.

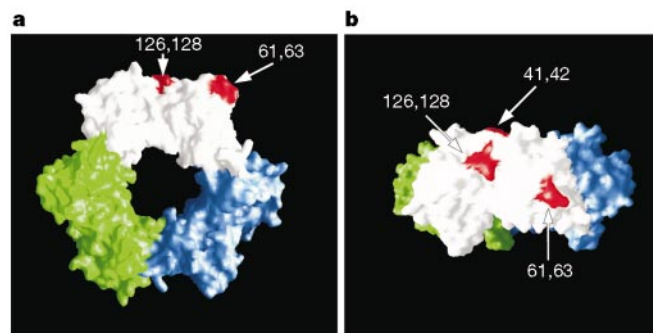


Figure 5 Molecular surface of the PCNA trimer. **a**, End view of the PCNA trimer. Subunits are coloured white, blue and green, respectively. Coloured in red are residues shown to affect silencing when mutated. **b**, Top view of the PCNA trimer. The PCNA trimer was rotated 90° along the x-axis.

Two classes of *pol30* mutants were discovered, on the basis of their combined effect with *cac1Δ* on TPE and *HMR* silencing, their effect on *HMR* silencing, and their location on the three-dimensional structure of PCNA. Genetic analysis suggests that PCNA is involved in silencing both dependently and independently of CAF-1, providing direct mechanisms to link inheritance of DNA to the propagation of epigenetic states of gene expression in eukaryotes, a process believed to determine cell fate in complex organisms³⁰. □

Methods

Strains, plasmids and other reagents

All of the yeast strains used in this study are listed in Table 2 of the Supplementary Information and were derived from the parental strain W303-1. Standard yeast media were used. Strains with deletion of the endogenous *POL30* gene originated from PKY741 (W303, *MATa pol30Δ::hisG-URA3-hisG*, pBL230(*POL30*)) from P. Kaufman. The *URA3* gene used to disrupt the *POL30* was deleted by selecting cells on plates containing the drug 5-fluoro-orotic acid (FOA). The plasmid pBL-211 (*URA3 ARS1 CEN4 POL30* in pBR322) from P. Burgers was used to replace pBL-230¹⁸. The resulting strain ZGY002 was used to cross with other strains to generate strains lacking the endogenous *POL30* gene. Sixteen *pol30* mutant plasmids were also provided by P. Burgers.

Detailed procedures regarding C-terminal tagging of Cac1 and Cac2, yeast transformation and plasmid shuffle, cloning and Southern blotting analysis of telomere length, sensitivity of yeast cells towards DNA damage agent, and production of antiserum against PCNA are described in detail in the Supplementary Information.

Assays for silencing

To assay telomeric silencing, the *URA3* gene integrated at the telomere of the left arm of chromosome VII was used¹² and procedures described⁹ were followed. Briefly, freshly grown yeast cells were taken from YPD plates, and diluted to an optical density at 600 nm (OD_{600}) of 0.6. Then tenfold serial dilutions were made, and 5-μl cells from each of the dilutions were spotted onto YPD, SCM-URA, and SCM+FOA plates (SCM, synthetic complete medium). During the initial screen, two independent clones for each of the *pol30* mutants were analysed and the results from the two clones agreed well with each other. To determine quantitatively the proportion of silenced cells in a population, cells grown to early logarithmic phase at 30 °C were diluted to an OD_{600} value of 0.7. Then tenfold serial dilutions were made, and 100-μl cells from each dilution were plated onto a FOA plate or a YPD plate. After two days of growth at 30 °C, cells on YPD plates were counted to assess viability. Cells on FOA plates were counted after eight days of incubation at 30 °C. To assay silencing of *HMR*, we used the *ADE2* gene to replace *HMR-a* gene to monitor changes in colony colour¹³. About 100 yeast cells from wild-type PCNA or each of the mutants were spread onto YPD plates. The plates were incubated at 30 °C for three days, and photos were taken after additional incubation at 4 °C for at least three days. To test the stability of light pink (repressed) or white (de-repressed) cells of the *pol30-8* mutant, a pink or white colony was taken from a YPD plate and re-suspended into water. About 100 cells were immediately plated onto a YPD plate (time 0) to assay expression of the *ADE2* gene, or grown in liquid YPD medium for a specific time before assaying the expression of the *ADE2* gene as described above. If over 50% of the size of a colony is white, we counted it as a white colony. A light pink colony of *pol30-8* has an appearance almost indistinguishable from that of a wild-type colony except that it is a little lighter in colour. A sectoring colony appears to have strips of white and pink cells (see Fig. 2b).

Immunoprecipitation of Cac2-13Myc

Yeast cells with Cac1 or without Cac1 were grown to similar OD and harvested. Yeast whole cell extracts were prepared by bead beating and clarified by centrifugation. The resulting lysates were incubated overnight at 4 °C with monoclonal antibody against the Myc epitope (9E10) cross-linked to protein G sepharose. After washing unbound proteins, the bound proteins were eluted using SDS sample buffer, resolved by 12.5% SDS-polyacrylamide gel electrophoresis (SDS-PAGE) and subjected to Western blot analysis.

In vitro glutathione S-transferase pull-down assay

Wild-type PCNA or mutant proteins were produced using the TNT T7 Quick coupled reticulocyte lysate system (Promega). For a 50-μl *in vitro* translation, 0.5 μg of DNA from wild type or each of the mutants cloned into pET3c were used. To test whether wild-type PCNA bound GST-Cac1, 20 μl of glutathione sepharose bound with GST-Cac1 or GST-REGα were incubated with 10 μl of *in vitro* translated ³⁵S-methionine-labelled PCNA in buffer D (25 mM Tris, pH 7.2, 100 mM NaCl) at an incubation volume of 100 μl or 500 μl. After 6 h incubation at 4 °C, the beads were washed four times, each with 0.5 ml of buffer D. Proteins that bound the beads were eluted with 20 μl of SDS sample buffer and were resolved on a 12.5% SDS-PAGE, and subjected to autoradiography.

To test whether PCNA mutants bound GST-Cac1, two different amounts of GST-Cac1 were incubated with an equal amount of *in vitro* translated PCNA or each of the mutants for 6 h in a 500-μl reaction. After washing the beads, all of the bound proteins were analysed as described above.

In vivo chromatin binding assay

Yeast cells with wild-type PCNA or each of the PCNA mutants indicated in the figure legends were grown in 5 ml of YPD medium at 25 °C overnight. Each culture was diluted to

10 ml YPD to an OD₆₀₀ value of 0.2 except *pol30-6* and *pol30-42*, which were diluted to an OD₆₀₀ value of 0.25, as these two mutants grew more slowly. When the OD₆₀₀ value of each culture reached 0.8–1, 0.5 ml of each culture was used for FACS analysis as described¹⁵. The rest of the yeast cells were harvested, and washed with 20% cold glycerol plus 2 mM of the protease inhibitor pefabloc. The chromatin-binding assay was performed as described¹⁵ except that twice the amounts of PMSF and pefabloc were used for all the buffers containing these two inhibitors. Equivalent amounts of each sample were then loaded onto 12.5% SDS–PAGE, transferred to nitrocellulose membranes, and probed with monoclonal antibodies against Cac1-3HA (12CA5, 1:10000), Cac2-13Myc (9E10, 1:10000), Orc3 (SB3, 1:10000) and polyclonal antisera to PCNA (871, 1:2000).

Received 19 April; accepted 19 September 2000.

1. Loo, S. & Rine, J. Silencing and heritable domains of gene expression. *Annu. Rev. Cell Dev. Biol.* **11**, 519–548 (1995).
2. Grunstein, M. Yeast heterochromatin: regulation of its assembly and inheritance by histones. *Cell* **93**, 325–328 (1998).
3. Lustig, A. J. Mechanisms of silencing in *Saccharomyces cerevisiae*. *Curr. Opin. Genet. Dev.* **8**, 233–239 (1998); erratum *Curr. Opin. Genet. Dev.* **8**, 721 (1998).
4. Pillus, L. & Rine, J. Epigenetic inheritance of transcriptional states in *S. cerevisiae*. *Cell* **59**, 637–647 (1989).
5. Waga, S. & Stillman, B. The DNA replication fork in eukaryotic cells. *Annu. Rev. Biochem.* **67**, 721–751 (1998).
6. Verreault, A. De novo nucleosome assembly: new pieces in an old puzzle. *Genes Dev.* **14**, 1430–1438 (2000).
7. Shibahara, K. & Stillman, B. Replication-dependent marking of DNA by PCNA facilitates CAF-1-coupled inheritance of chromatin. *Cell* **96**, 575–585 (1999).
8. Moggs, J. G. *et al.* A CAF-1-PCNA-mediated chromatin assembly pathway triggered by sensing DNA damage. *Mol. Cell Biol.* **20**, 1206–1218 (2000).
9. Kaufman, P. D., Kobayashi, R. & Stillman, B. Ultraviolet radiation sensitivity and reduction of telomeric silencing in *Saccharomyces cerevisiae* cells lacking chromatin assembly factor-I. *Genes Dev.* **11**, 345–357 (1997).
10. Enomoto, S. & Berman, J. Chromatin assembly factor I contributes to the maintenance, but not the re-establishment, of silencing at the yeast silent mating loci. *Genes Dev.* **12**, 219–232 (1998).
11. Henderson, D. S., Banga, S. S., Grigliatti, T. A. & Boyd, J. B. Mutagen sensitivity and suppression of position-effect variegation result from mutations in *mus209*, the *Drosophila* gene encoding PCNA. *EMBO J.* **13**, 1450–1459 (1994).
12. Gottschling, D. E., Aparicio, O. M., Billington, B. L. & Zakian, V. A. Position effect at *S. cerevisiae* telomeres: reversible repression of Pol II transcription. *Cell* **63**, 751–762 (1990).
13. Sussel, L., Vannier, D. & Shore, D. Epigenetic switching of transcriptional states: cis- and trans-acting factors affecting establishment of silencing at the HMR locus in *Saccharomyces cerevisiae*. *Mol. Cell Biol.* **13**, 3919–3928 (1993).
14. Mahoney, D. J., Marquardt, R., Shei, G. J., Rose, A. B. & Broach, J. R. Mutations in the HML E silencer of *Saccharomyces cerevisiae* yield metastable inheritance of transcriptional repression. *Genes Dev.* **5**, 605–615 (1991).
15. Liang, C. & Stillman, B. Persistent initiation of DNA replication and chromatin-bound MCM proteins during the cell cycle in *cdc6* mutants. *Genes Dev.* **11**, 3375–3386 (1997).
16. Kaufman, P. D., Kobayashi, R., Kessler, N. & Stillman, B. The p150 and p60 subunits of chromatin assembly factor I: a molecular link between newly synthesized histones and DNA replication. *Cell* **81**, 1105–1114 (1995).
17. Krishna, T. S., Kong, X. P., Gary, S., Burgers, P. M. & Kuriyan, J. Crystal structure of the eukaryotic DNA polymerase processivity factor PCNA. *Cell* **79**, 1233–1243 (1994).
18. Ayyagari, R., Impellizzeri, K. J., Yoder, B. L., Gary, S. L. & Burgers, P. M. A mutational analysis of the yeast proliferating cell nuclear antigen indicates distinct roles in DNA replication and DNA repair. *Mol. Cell Biol.* **15**, 4420–4429 (1995).
19. Eissenberg, J. C., Ayyagari, R., Gomes, X. V. & Burgers, P. M. Mutations in yeast proliferating cell nuclear antigen define distinct sites for interaction with DNA polymerase delta and DNA polymerase epsilon. *Mol. Cell Biol.* **17**, 6367–6378 (1997).
20. Miller, A. M. & Nasmyth, K. A. Role of DNA replication in the repression of silent mating type loci in yeast. *Nature* **312**, 247–251 (1984).
21. Triolo, T. & Sternglanz, R. Role of interactions between the origin recognition complex and SIR1 in transcriptional silencing. *Nature* **381**, 251–253 (1996).
22. Chien, C. T., Buck, S., Sternglanz, R. & Shore, D. Targeting of SIR1 protein establishes transcriptional silencing at HM loci and telomeres in yeast. *Cell* **75**, 531–541 (1993).
23. Fox, C. A., Ehrenhofer-Murray, A. E., Loo, S. & Rine, J. The origin recognition complex, SIR1, and the S phase requirement for silencing. *Science* **276**, 1547–1551 (1997).
24. Murzina, N., Verreault, A., Laue, E. & Stillman, B. Heterochromatin dynamics in mouse cells: interaction between chromatin assembly factor 1 and HP1 proteins. *Mol. Cell* **4**, 529–540 (1999).
25. Le, S., Davis, C., Konopka, J. B. & Sternglanz, R. Two new S-phase-specific genes from *Saccharomyces cerevisiae*. *Yeast* **13**, 1029–1042 (1997).
26. Tyler, J. K. *et al.* The RCAF complex mediates chromatin assembly during DNA replication and repair. *Nature* **402**, 555–560 (1999).
27. Verreault, A., Kaufman, P. D., Kobayashi, R. & Stillman, B. Nucleosome assembly by a complex of CAF-1 and acetylated histones H3/H4. *Cell* **87**, 95–104 (1996).
28. Ehrenhofer-Murray, A. E., Kamakaka, R. T. & Rine, J. A role for the replication proteins PCNA, RF-C, polymerase epsilon and Cdc45 in transcriptional silencing in *Saccharomyces cerevisiae*. *Genetics* **153**, 1171–1182 (1999).
29. Laman, H., Balderes, D. & Shore, D. Disturbance of normal cell cycle progression enhances the establishment of transcriptional silencing in *Saccharomyces cerevisiae*. *Mol. Cell Biol.* **15**, 3608–3617 (1995).
30. Scott, M. P. Development: the natural history of genes. *Cell* **100**, 27–40 (2000).

Supplementary information is available at Nature's World-Wide Web site (<http://www.nature.com>) or as paper copy from the London editorial office of Nature.

Acknowledgements

We thank P. Burgers, P. Kaufman, R. Sternglanz and D. Shore for plasmids and yeast strains used in this study. We thank A. Stenlund for critical reading of the manuscript, T. Tully for statistical analysis of the data presented in Table 1, and members of the Stillman laboratory, especially L. Zou, for helpful discussions. This work is supported by a grant from the National Institutes of Health (to B. S.). Z. Z. is supported by a postdoctoral fellowship from the Cancer Research Fund of the Damon Runyon-Walter Winchell Foundation. K. S. is a Leukemia Society of America Special Fellow.

Correspondence and requests for materials should be addressed to B. S. (e-mail: stillman@cshl.org).

A transcription reinitiation intermediate that is stabilized by activator

Natalya Yudkovsky*, Jeffrey A. Ranish*† & Steven Hahn*‡

* Division of Basic Sciences, The Fred Hutchinson Cancer Research Center, and Molecular and Cellular Biology Program, University of Washington, Seattle, Washington 98109, USA

‡ The Howard Hughes Medical Institute, Seattle, Washington 98109, USA

High levels of gene transcription by RNA polymerase II depend on high rates of transcription initiation and reinitiation. Initiation requires recruitment of the complete transcription machinery to a promoter, a process facilitated by activators and chromatin remodelling factors. Reinitiation probably occurs through a different pathway¹. After initiation, a subset of the transcription machinery remains at the promoter, forming a platform for assembly of a second transcription complex^{2–4}. Here we describe the isolation of a reinitiation intermediate that includes transcription factors TFIID, TFIIA, TFIIF, TFIIE and Mediator. This intermediate can act as a scaffold for formation of a functional reinitiation complex. Formation of this scaffold is dependent on ATP and TFIIF. The scaffold is stabilized in the presence of the activator Gal4–VP16, but not Gal4–AH, suggesting a new role for some activators and Mediator in promoting high levels of transcription.

The first step in transcription initiation by RNA polymerase II (RNA Pol II) is recruitment of the transcription machinery to a promoter to form a pre-initiation complex (PIC). After initiation, a subset of the factors in the PIC dissociates from the promoter^{2–4}. To begin a second round of transcription (reinitiation), this subset of factors, along with RNA Pol II, must again be recruited to the promoter. In the yeast Mediator-dependent system, PIC formation can occur in at least two steps⁵. In the first step, TFIID and TFIIA bind cooperatively to the promoter. In the second step, the rest of the transcription machinery stably binds to form a complete PIC. This step requires cooperative binding of TFIIB and holopolymerase, a complex composed of RNA Pol II and Mediator^{6,7}. Mediator, which contains Srb, Med and other proteins, associates with RNA Pol II, allows transcription to be responsive to activators, and stimulates RNA Pol II carboxy-terminal domain (CTD) phosphorylation^{6,7}. Although reinitiation probably involves the same complement of transcription factors as initiation, evidence suggests that the pathways for initiation and reinitiation are distinct. *In vitro* studies using HeLa extracts have shown that the rate of reinitiation at some promoters is fourfold higher than that of initiation⁸. Other studies showed that activator, TFIID and TFIIA remain at the promoter after initiation^{2–4}.

We used an immobilized promoter template assay and yeast nuclear extracts to isolate a reinitiation intermediate⁵. Such an

† Present address: Institute for Systems Biology, Seattle, Washington 98105, USA.

approach allowed us to monitor the entire transcription machinery in a crude Mediator-dependent transcription system, rather than one using purified factors. In this assay, HIS4 promoter templates immobilized on magnetic beads were incubated with nuclear extract and the activator Gal4-AH to allow PIC formation. The PICs were then washed and transcription initiated by addition of nucleotides for 2 min. This procedure allows only a single round of transcription to occur, as the transcription signal detected is equivalent to that seen after incubation of PICs with NTPs for 1 min, followed by addition of sarkosyl to block reinitiation (data not shown). After nucleotide addition, proteins still bound to the templates were isolated. As expected from previous studies²⁻⁴, RNA Pol II, TFIIB and TFIIF dissociated from the templates, but activator, TBP, the TFIID subunits TAF_{II}90 and TAF_{II}67 and TFIIA remained bound to the promoters (Fig. 1, compare lanes 2 and 3; and data not shown). We found that the Mediator complex (Srb4, Srb2, Med6 and Gal11 subunits) and substantial amounts of TFIIF and TFIIE also remained at the promoter. Specifically, the level of RNA Pol II was reduced 13-fold, the level of TFIIB was

reduced 24-fold and the level of TFIIF was reduced 14-fold, whereas the levels of all other components were reduced less than 2.5-fold.

We then investigated whether this complex of activator, TFIID, TFIIA, TFIIF, TFIIE and Mediator could function as a reinitiation intermediate, by acting as a scaffold on which a functional transcription complex would reassemble. The scaffolds were formed and washed as described in Fig. 1. A second nuclear extract was then added along with nucleotides to determine whether a second round of transcription could occur (Fig. 2a). For the second extract, we used extracts made from strains with mutations in Mediator components (Δ Srb2 or Srb4ts), TFIIB (G41E), TFIIF (Kin28ts), TFIIE (Tfa1ts), TBP (I143N) or TFIIA (Toa1-25). All of these extracts are defective in PIC assembly⁵ (data not shown) and transcription (Fig. 2b, lanes 1-7; Fig. 2c, lanes 1-3; Fig. 2d, lanes 1-5). As a control to show that few active PICs remained after the first round of transcription, very little RNA was produced when nucleotides were added to the scaffolds in the absence of a second extract (Fig. 2b, lane 8; Fig. 2c, lane 4; Fig. 2d, lane 6).

When supplemented with extracts from Δ Srb2, Srb4ts, Kin28ts, TBP143N and Toa1-25 mutants, the scaffolds supported a second round of transcription, confirming the presence of these components

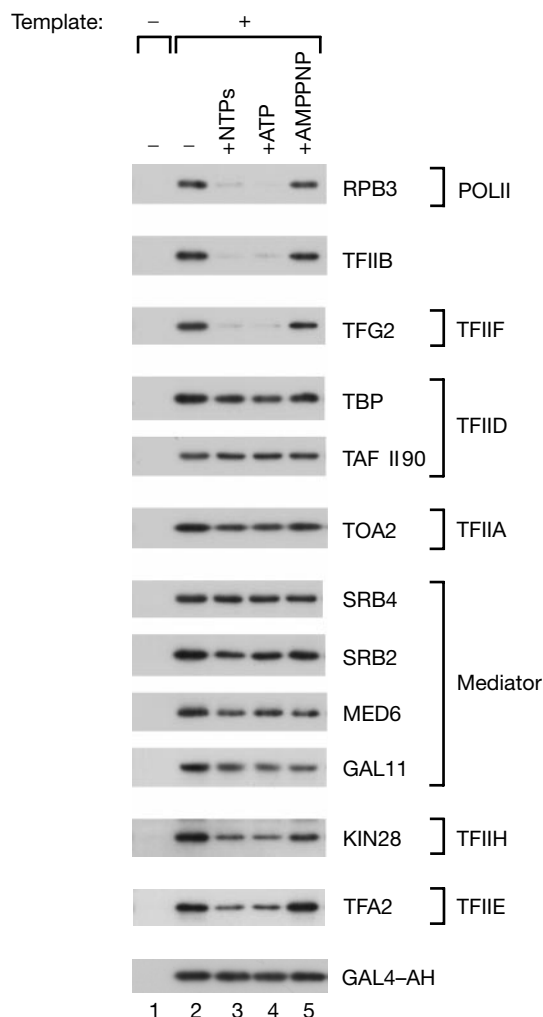


Figure 1 Scaffold contains activator, TFIID, TFIIA, Mediator, TFIIF and TFIIE. The 515 template, comprising a HIS4 promoter with a single Gal4 DNA-binding site upstream, was immobilized on a magnetic Dynabead. Immobilized templates were incubated with the activator Gal4-AH and nuclear extract for 40 min to form PICs. Templates were washed and nucleotides added for 2 min as indicated. Templates were washed again, and bound proteins were isolated by *Pst*I digestion and detected by western blotting. Lane 2 shows a typical PIC. As a control for nonspecific binding to the Dynabeads, the reaction in lane 1 was performed without template.

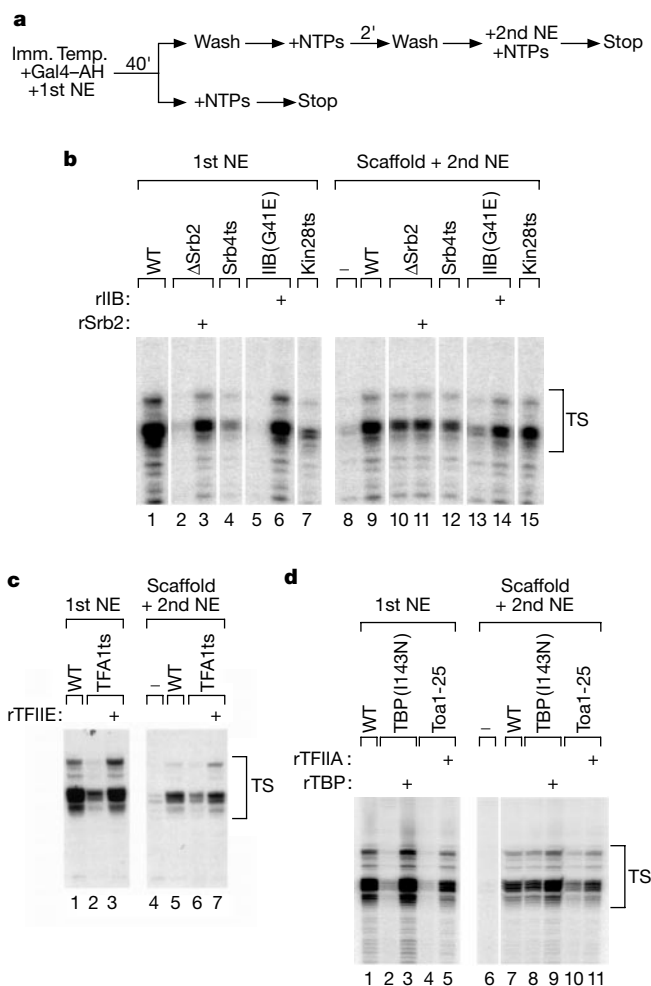


Figure 2 Scaffold supports reinitiation. **a**, The scaffold reinitiation assay. **b**, Lanes 1-7, nuclear extracts were incubated with 515 immobilized templates for 40 min to form PICs. NTPs were added and reactions stopped after 2 min to allow for a single round of transcription. Lanes 8-15, reactions were performed as described in **a**. NTPs were added along with the second nuclear extract for 2 min. As a control for residual active complexes, no second nuclear extract was added in lane 8. **c**, **d**, Reactions were performed as described in **b**, but in **d** all final NTP incubations occurred for 30 min. Reactions were assayed by primer extension. TS, transcription signal. WT, wild type.

in a functional reinitiation intermediate (Fig. 2b, lanes 8–12, 15; Fig. 2d, lanes 6–11). Little transcription was seen with the IIB(G41E) extract, however, confirming that TFIIB is not part of the scaffold (Fig. 2b, lanes 13 and 14). Notably, although recombinant Srb2 restored transcriptional activity to the Δ Srb2 extract, it had no effect on transcription when the scaffold was used (Fig. 2b, compare lanes 2 and 3 with 10 and 11). Although recombinant TBP and TFIIA did stimulate transcription on the scaffold when the TBPI143N and Toal-25 extracts were used, transcription in their absence was significantly higher than that seen in the absence of scaffold (Fig. 2d, compare lanes 2 and 4 with 8 and 10). Recombinant TFIIE stimulated transcription from the Tfa1ts extract fourfold (Fig. 2c, lanes 2 and 3), as compared with twofold stimulation when scaffold templates were used (lanes 6 and 7). From this and from Fig. 1, we conclude that the scaffold contains some functional TFIIE. It is apparent that TFIIE is the least stable component of the scaffold, and that TFIIH and TFIIA also dissociate to some extent on NTP addition. As a control for the experiments of Fig. 2, we used a competition assay to show that most of the transcription observed in the second round of initiation originated from scaffolds, rather than from newly formed PICs. As expected, when a second template was added to scaffolds along with the second nuclear extract, most of the transcription observed originated from the scaffold template (see Supplementary Information).

We were able to isolate this functional reinitiation intermediate despite the low percentage of active PICs in our assay. We determined the number of active PICs by measuring the amount of RNA produced in a single round of transcription. Comparing this number to the total number of PICs formed showed that only 5–10% of PICs were active in transcription (data not shown). These data indicate that the scaffold complexes isolated by our assay may be the result of dissociation of both active and inactive PICs. As these complexes can support reinitiation, these results imply that both active and inactive PICs dissociate by the same mechanism on nucleotide addition.

We found that adding only ATP to PICs had the same effect as

adding all four nucleotides: both resulted in PIC dissociation and loss of active PICs (Fig. 1; and data not shown). The ATP analogue AMPPNP did not promote PIC dissociation (Fig. 1, lane 5), which suggests that ATP hydrolysis, rather than transcription, is necessary for PIC dissociation. We therefore attempted to identify a PIC component with ATP-dependent activity that might be responsible for PIC dissociation. Three subunits of TFIIH were good candidates: the helicases Rad25 and Rad3, and the CTD kinase Kin28. We prepared nuclear extracts from transcriptionally defective strains that contained temperature-sensitive mutations in either Rad3 or Kin28 and tested them for PIC dissociation. These extracts were able to form a PIC intermediate that lacked both the Kin28 and Tfb1 subunits of TFIIH, even though these subunits were present in the mutant extracts (Fig. 3, lanes 2, 4, 6; and data not shown). The lack of these subunits suggests that this PIC intermediate lacks the entire TFIIH complex. After ATP addition, PICs lacking TFIIH were not able to dissociate into scaffolds, indicating that PIC dissociation is dependent on ATP and TFIIH (Fig. 3, lanes 3, 5, 7). Because PICs formed with both TFIIH mutant extracts probably lacked the entire TFIIH complex, we were unable to determine which of the TFIIH subunits is necessary for PIC dissociation. Studies have shown that phosphorylated, elongating RNA Pol II is not associated with Mediator⁹, suggesting that CTD phosphorylation by Kin28 may be required for the dissociation of Mediator from RNA Pol II during scaffold formation.

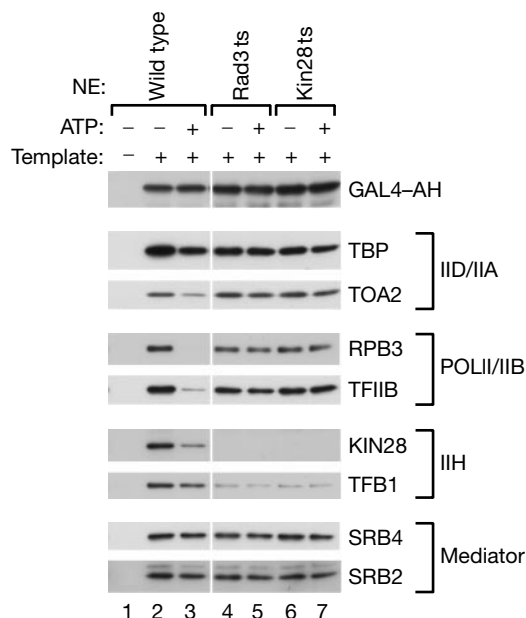


Figure 3 Scaffold formation is TFIIH-dependent. The immobilized template assay was performed as described in Fig. 1, using the indicated nuclear extracts. ATP was added to the reactions where indicated. Factors bound to the templates were assayed by SDS-PAGE and western blot. A typical wild-type PIC and a typical wild-type scaffold are shown in lanes 2 and 3, respectively. As a control for nonspecific binding to Dynabeads, the reaction in lane 1 was performed without template.

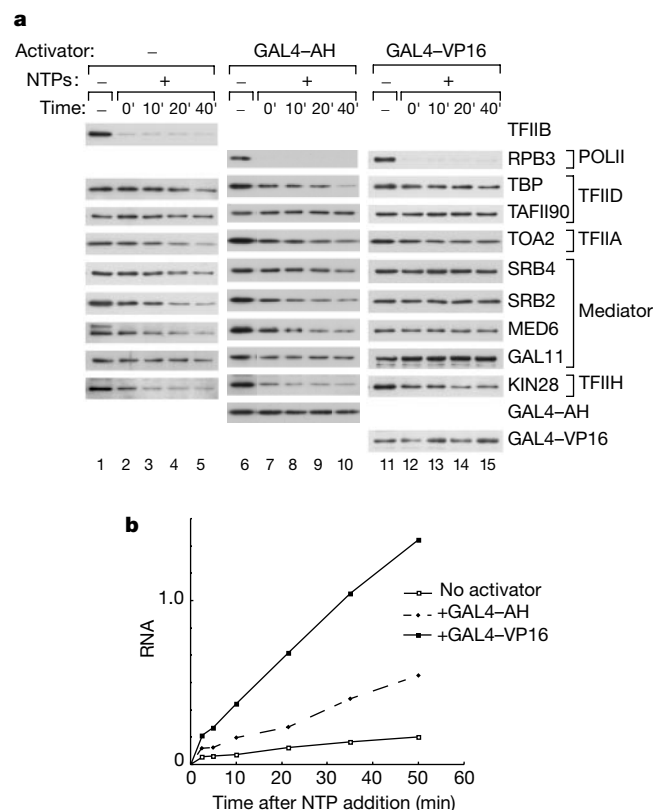


Figure 4 Gal4-VP16 promotes scaffold stability and a higher rate of reinitiation. **a**, 515 immobilized templates were pre-incubated with Gal4-AH, Gal4-VP16 or no activator. Scaffolds were formed using wild-type nuclear extract as described in Fig. 1. Scaffolds were incubated in transcription buffer for the times indicated, washed for 1 min, and the bound proteins analysed by western blot. As controls, PICs are shown in lanes 1, 6 and 11. **b**, Wild-type nuclear extract was incubated with pSH515, a plasmid containing the HIS4 template (Fig. 1) and Gal4-AH, Gal4-VP16 or no activator to form PICs. NTPs were added and samples removed for primer extension at 2.5, 5, 10, 21.5, 35 and 50 min. The transcription signal is shown plotted against time. The 2.5 min time point is equivalent to a single round of transcription.

We also analysed the effects of activator on scaffold formation and stability. Although activators function by stimulating PIC formation through transcription factor recruitment, evidence indicates that they also have a role in reinitiation. The heat shock factor⁴ and oestrogen receptor¹⁰ transcription activators, as well as the HIV-1 enhancer¹¹, stimulate reinitiation *in vitro*. Other *in vitro*² and *in vivo*¹² experiments showed that the presence of activator at the promoter is required for continued high levels of transcription. As PIC dissociation can occur in the absence of activator (Fig. 4a, lanes 1 and 2), we measured scaffold stability in either the absence of activator, or the presence of the activators Gal4-AH or Gal4-VP16. The scaffold was formed as described in Fig. 1 and analysed by western blot after incubation in transcription buffer for up to 40 min (Fig. 4a). In either the absence of activator or the presence of Gal4-AH, the levels of TBP, TFIIA, Srb4, Srb2 and Med6 decreased by 3–5-fold after 40 min. In contrast, when Gal4-VP16 was used the levels of all of these factors remained steady after 40 min. These results show that although the scaffold reinitiation intermediate can be formed without activator it is more stable in the presence of Gal4-VP16. As some activators can interact with TFIID¹³, TFIIA¹⁴ and various Mediator components^{15,16}, this stabilization is probably due to interactions between activators and scaffold components.

We next measured the rate of multiround transcription in the absence of activator and in the presence of Gal4-AH or Gal4-VP16. Nucleotides were added to PICs formed with wild-type nuclear extract, and transcription was measured at various time points (Fig. 4b). When the transcription signal is plotted against time, the resulting curve shows biphasic kinetics of transcription^{1,17}. RNA is rapidly produced from the preformed PICs, followed by a slower rate of RNA synthesis resulting from reinitiation and new initiation events. As, in the presence of activators, the rate of RNA synthesis after the initial burst of transcription is much faster than the rate of initial PIC formation, most of the subsequent RNA synthesis probably results from reinitiation rather than new initiation

events¹⁷. We found that with Gal4-VP16, the rate of transcription after the first round was 10-fold higher than with no activator and 3-fold higher than with Gal4-AH. These data show a correlation between scaffold stability and the rate of reinitiation, and support a role in scaffold stability for some activators in reinitiation.

Our results suggest a model for reinitiation in which activator, Mediator, TFIID, TFIIA, TFIIF and TFIIE remain at the promoter after RNA Pol II initiates transcription (Fig. 5). These factors are components of a scaffold on which other factors can assemble to form a reinitiation complex. As the binding of TFIID to promoters is a rate-limiting step in transcription initiation *in vivo*^{18,19}, such a model can account for the observation that rates of reinitiation are higher than those of initiation⁸. In this model, the activator could play a dual role in promoting high levels of reinitiation. First, after initiation, activator could directly promote the recruitment of the missing components of the transcription machinery. Second, some activators such as VP-16 can directly stabilize the scaffold complex to promote reinitiation. Our model is supported by *in vivo* artificial recruitment assays in which high levels of transcription are achieved by fusing scaffold components to DNA binding domains^{19–21}. Although these high levels of transcription have been interpreted as resulting from an increase in factor recruitment, our model suggests that they could also result from an increase in reinitiation owing to greater scaffold stability. This model predicts that the rate of transcription initiation on a scaffold template would be higher than that on a naked template. Indeed, our experiments have shown that the presence of scaffold stimulates initial rates of transcription 2–3-fold (see Supplementary Information), consistent with results showing that previously transcribed templates are preferentially transcribed²². Our model also suggests that holopolymerase is not involved in reinitiation. Although there is evidence that Mediator subunits are responsible for transcription of most genes in yeast²³, the presence of Mediator in the scaffold suggests that holopolymerase participates only in initiation and not in reinitiation. Instead, reinitiation may involve the recruitment of free RNA Pol II, or RNA Pol II in a distinct complex, to the scaffold. □

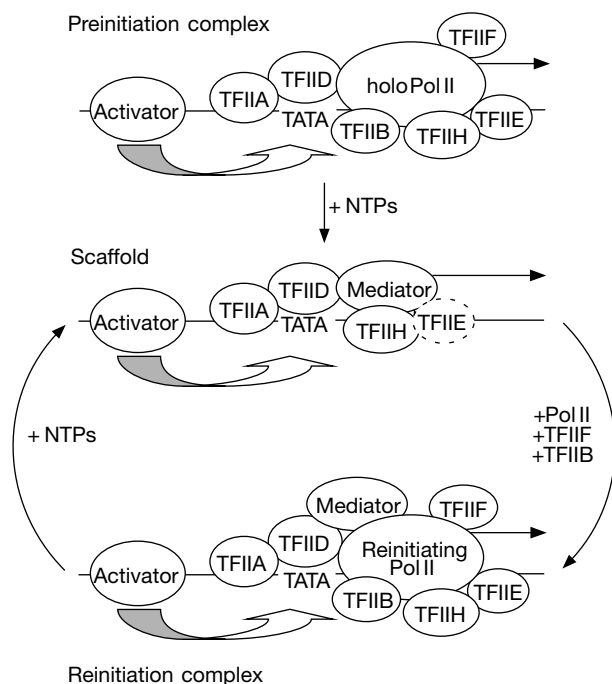


Figure 5 Reinitiation model. When NTPs are added to a pre-initiation complex, RNA Pol II initiates transcription. Whereas TFIIF and TFIIE dissociate from the promoter, activator, TFIID, TFIIA, TFIIF, TFIIE and Mediator are left behind in a scaffold complex. RNA Pol II, TFIIF and TFIIE then reassemble onto the scaffold to form a complex capable of reinitiating transcription. TFIIE is shown as being the least stable scaffold component.

Methods

Preparation of nuclear extracts

All yeast strains have been described⁵, except for Rad3-ts14 (ref. 24) and Kin28-ts16 (ref. 25). Nuclear extracts were prepared as described previously²⁶ and on the World-Wide Web (www.fhcr.org/labs/hahn).

Immobilized template assay

Immobilized templates were prepared as described⁵. PIC formation experiments were performed as described on the World-Wide Web (www.fhcr.org/labs/hahn). The reactions were run on a 4–12% NuPAGE gel (NOVEX) and transferred to Immobilon membranes (Millipore). Proteins were detected using Pierce ECL kits. Band intensities were determined by densitometry using IQMACv1.2 software (Molecular Dynamics). Scaffold isolation was performed similarly, except that after washing, PICs were resuspended in 100 µl transcription mix, and incubated with 1 µg HaeIII-digested *Escherichia coli* DNA competitor, and 100 µM NTPs, ATP or AMPPNP for 2 min at room temperature. The templates were washed once with wash buffer, isolated by digestion with 60 units PstI for 30 min at 37 °C, and processed as described above. For the scaffold stability experiment, scaffolds were isolated as described above, except that after being washed, they were resuspended in transcription mix with 1 µg HaeIII-digested *E. coli* DNA competitor. Aliquots of 100 µl were removed at the indicated times, washed once with wash buffer, and isolated and processed as described above.

Transcription

We carried out plasmid transcription by incubating wild-type nuclear extract with the HIS4-promoter-containing plasmid, pSH515 or pSH559 as described previously²⁷ and on the World-Wide Web (www.fhcr.org/labs/hahn). pSH559 was made by digesting pSH515 with BamHI and SfoI to delete 50 base pairs of transcribed sequence. In experiments where both pSH515 and pSH559 were used, the Cyc1 primer (5'-GAGAGGCGGTTTTCGTAT TGGG-3') was used for primer extension. We carried out transcription on immobilized templates as described⁵. The RNA was isolated by phenol:chloroform (2:1) extraction and ethanol precipitation. Primer extension was performed on the RNA samples as described using either the LacI primer or the Cyc1 primer²⁷. For scaffold functional assays, scaffolds were formed as described above using wild-type nuclear extracts. After washing, scaffolds were resuspended in transcription mix containing 120–180 µg of a second nuclear extract

and 500 ng *Hae*II-digested *E. coli* DNA competitor. NTPs were added to 100 μ M either immediately, or after a 40-min incubation. The reactions were stopped after either 2 or 30 min, and analysed by primer extension. For the scaffold competition experiment, scaffolds were formed on the 559 immobilized template as described above, using Δ Srb2 nuclear extract with 100 ng rSrb2. After washing, an equivalent amount of 515 immobilized competitor template was added, and the reactions were resuspended in transcription mix containing Δ Srb2 nuclear extract either with or without 100 ng rSrb2. Reactions were incubated for either 10 or 20 min at room temperature. NTPs were then added for 2 min, and the reactions were stopped and analysed by primer extension using the Cyc1 primer. All transcription signals were quantified by PhosphorImager (Molecular Dynamics).

PIC and RNA quantification

The total number of PICs formed in an immobilized template assay was quantified by a western blot comparison with known amounts of purified recombinant TBP, TFIIA, TFIIB and TFIIE. Band intensities were determined by densitometry using IQMACv1.2 software (Molecular Dynamics). We determined the number of active PICs by quantifying the amount of RNA produced in a single round of transcription. The amount of RNA produced in a single round of immobilized template transcription was assayed by S1 nuclease protection using a 5'-end-labelled DNA oligonucleotide probe (5'-GTAACT ATTGATATTACTATTACACAGCGCAGGGTGTAG-3'). S1 nuclease protection using the same probe was also performed on increasing quantities of a 30-nucleotide RNA oligonucleotide standard to quantify the amount of RNA produced in the transcription reaction. This amount was taken to be equivalent to the number of active PICs.

Received 3 August; accepted 25 September 2000.

- Hahn, S. Activation and the role of reinitiation in the control of transcription by RNA polymerase II. *Cold Spring Harb. Symp. Quant. Biol.* **63**, 181–188 (1998).
- Roberts, S. G., Choy, B., Walker, S. S., Lin, Y. S. & Green, M. R. A role for activator-mediated TFIIB recruitment in diverse aspects of transcriptional regulation. *Curr. Biol.* **5**, 508–516 (1995).
- Zawel, L., Kuman, K. P. & Reinberg, D. Recycling of the general transcription factors during RNA polymerase II transcription. *Genes Dev.* **9**, 1479–1490 (1995).
- Sandaltzopoulos, R. & Becker, P. B. Heat shock factor increases the reinitiation rate from potentiated chromatin templates. *Mol. Cell. Biol.* **18**, 361–367 (1998).
- Ranish, J. A., Yudkovsky, N. & Hahn, S. Intermediates in formation and activity of the RNA polymerase II preinitiation complex: holoenzyme recruitment and a postrecruitment role for the TATA box and TFIIB. *Genes Dev.* **13**, 49–63 (1999).
- Kim, Y. J., Bjorklund, S., Li, Y., Sayre, M. H. & Kornberg, R. D. A multiprotein mediator of transcriptional activation and its interaction with the C-terminal repeat domain of RNA Polymerase II. *Cell* **77**, 599–608 (1994).
- Koleske, A. J. & Young, R. A. An RNA polymerase II holoenzyme responsive to activators. *Nature* **368**, 466–469 (1994).
- Jiang, Y. & Gralla, J. D. Uncoupling of initiation and reinitiation rates during HeLa RNA polymerase II transcription in vitro. *Mol. Cell. Biol.* **13**, 4572–4577 (1993).
- Svejstrup, J. Q. *et al.* Evidence for a mediator cycle at the initiation of transcription. *Proc. Natl Acad. Sci. USA* **94**, 6075–6078 (1997).
- Kraus, W. L. & Kadonaga, J. T. p300 and estrogen receptor cooperatively activate transcription via differential enhancement of initiation and reinitiation. *Genes Dev.* **12**, 331–342 (1998).
- Sheridan, P. L., Mayall, T. P., Verdin, E. & Jones, K. A. Histone acetyltransferases regulate HIV-1

- enhancer activity in vitro. *Genes Dev.* **11**, 3327–3340 (1997).
- Ho, S. N., Biggar, S. R., Spencer, D. M., Schreiber, S. L. & Crabtree, G. R. Dimeric ligands define a role for transcriptional activation domains in reinitiation. *Nature* **382**, 822–826 (1996).
- Stringer, K. F., Ingles, C. J. & Greenblatt, J. Direct and selective binding of an acidic transcriptional activation domain to the TATA-box factor TFIID. *Nature* **345**, 783–786 (1990).
- Ozer, J. *et al.* Molecular cloning of the small (γ) subunit of human TFIIA reveals functions critical for activated transcription. *Genes Dev.* **8**, 2324–2335 (1994).
- Koh, S. S., Ansari, A. Z., Ptashne, M. & Young, R. A. An activator target in the RNA polymerase II holoenzyme. *Mol. Cell* **1**, 895–904 (1998).
- Lee, Y. C., Park, J. M., Min, S., Han, S. J. & Kim, Y. J. An activator binding module of yeast RNA polymerase II holoenzyme. *Mol. Cell. Biol.* **19**, 2967–2976 (1999).
- Yean, D. & Gralla, J. Transcription reinitiation rate: a special role for the TATA box. *Mol. Cell. Biol.* **17**, 3809–3816 (1997).
- Klein, C. & Struhl, K. Increased recruitment of TATA-binding protein to the promoter by transcriptional activation domains in vivo. *Science* **266**, 280–282 (1994).
- Xiao, H., Friesen, J. D. & Lis, J. T. Recruiting TATA-binding protein to a promoter: transcription activation without an upstream activator. *Mol. Cell. Biol.* **15**, 5757–5761 (1995).
- Farrell, S., Simkovich, N., Wu, Y., Barberis, A. & Ptashne, M. Gene activation by recruitment of the RNA polymerase II holoenzyme. *Genes Dev.* **10**, 2359–2367 (1996).
- Keaveney, M. & Struhl, K. Activator-mediated recruitment of the RNA polymerase II machinery is the predominant mechanism for transcriptional activation in yeast. *Mol. Cell* **1**, 917–924 (1998).
- Hawley, D. K. & Roeder, R. G. Functional steps in transcription initiation and reinitiation from the major late promoter in a HeLa nuclear extract. *J. Biol. Chem.* **262**, 3452–3461 (1987).
- Holstege, F. C. *et al.* Dissecting the regulatory circuitry of a eukaryotic genome. *Cell* **95**, 717–728 (1998).
- Guzder, S. N. *et al.* DNA repair gene RAD3 of *S. cerevisiae* is essential for transcription by RNA polymerase II. *Nature* **367**, 91–94 (1994).
- Cismowski, M. J., Laff, G. M., Solomon, M. J. & Reed, S. I. KIN28 encodes a C-terminal domain kinase that controls mRNA transcription in *S. cerevisiae* but lacks cyclin-dependent kinase-activating kinase (CAK) activity. *Mol. Cell. Biol.* **15**, 2983–2992 (1995).
- Kang, J. J., Auble, D. T., Ranish, J. A. & Hahn, S. Analysis of the yeast transcription factor TFIIA: distinct functional regions and a polymerase II-specific role in basal and activated transcription. *Mol. Cell Biol.* **15**, 1234–1243 (1995).
- Ranish, J. A. & Hahn, S. The yeast general transcription factor TFIIA is composed of two polypeptide subunits. *J. Biol. Chem.* **266**, 19320–19327 (1991).

Supplementary information is available at Nature's World-Wide Web site (<http://www.nature.com>) or as paper copy from the London editorial office of Nature.

Acknowledgements

We thank members of the Hahn and Reeder laboratories for helpful discussions, and A. Krumm, S. Parkhurst and R. Reeder for comments on the manuscript. We also thank L. Prakash for providing the Rad3ts strain, M. Solomon for providing the Kin28ts strain, D. Reinberg for TFIIE antibodies and H. Sakurai for providing the Tfaltts strain and antibodies to Gal11. This work was supported by grants from the NIH to S.H. and an NIH training grant to N.Y. S.H. is an associate investigator of the Howard Hughes Medical Institute.

Correspondence and requests for materials should be addressed to S.H. (e-mail: shahn@fhcrc.org).

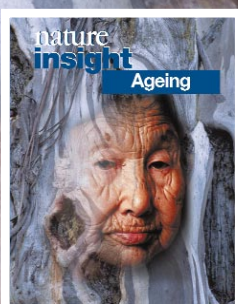


nature insight

Ageing

Cover illustrations

courtesy of P. O'Hara/
P. A. Souders/CORBIS.



Evolutionary history has determined that individuals thrive for long enough to produce and nurture their offspring. Thereafter, the ageing process involves a slow decline in physiological vigour and an increasing susceptibility to age-related disease. Much of human culture and thinking has been shaped by the inevitability of our ageing and death.

The current scientific picture of ageing presents us with an intriguing jigsaw puzzle. We know, for example, that reducing food intake can slow the ageing process, at least in lower animals such as nematode worms. We know that telomeres, which protect the ends of chromosomes, erode as our cells age. But how can we connect together these and other discoveries to give a meaningful picture of the genetic and biochemical processes that underlie ageing? With this goal in mind we have assembled a collection of review articles which rehearse our current understanding of the ageing process from several distinct vantage points.

Why do humans age, when fairly similar creatures (such as turtles) apparently do not? Kirkwood and Austad begin on page 233 by introducing the conundrum of ageing from an evolutionary standpoint. Finkel and Holbrook then review the role that oxidative damage plays in the ageing process, on page 239. Cancer is one of the most prevalent age-linked diseases in our society, and on page 248 DePinho discusses the ways in which age-related cancer may develop. Human ageing is a slow process, and for this reason researchers have often made productive use of organisms such as fruitflies, which have short life spans and can be manipulated in large numbers. Guarente and Kenyon review recent research on ageing in model organisms on page 255. Martin and Oshima, on page 263, then summarize our understanding of human 'progeroid' syndromes, in which elements of the normal ageing programme are accelerated owing to genetic abnormalities. In a concluding commentary article on page 267, Hayflick ponders the future prospects for research on ageing, and what implications the fruits of this research may have for our society.

Although some readers may feel that they are already more familiar with the ageing process than they might wish to be, we hope that this collection of reviews will nonetheless prove stimulating and informative.

Bernd Pulverer Associate Editor

Richard Turner Senior Editor

Ritu Dhand Insight Programme Editor

Liz Allen Publisher

Why do we age?

Thomas B. L. Kirkwood* & Steven N. Austad†

*Department of Gerontology, University of Newcastle, Institute for the Health of the Elderly, Newcastle General Hospital, Newcastle upon Tyne NE4 6BE, UK

†Department of Biological Sciences, University of Idaho, Moscow, Idaho 83844-3051, USA

The evolutionary theory of ageing explains why ageing occurs, giving valuable insight into the mechanisms underlying the complex cellular and molecular changes that contribute to senescence. Such understanding also helps to clarify how the genome shapes the ageing process, thereby aiding the study of the genetic factors that influence longevity and age-associated diseases.

Ageing is usually defined as the progressive loss of function accompanied by decreasing fertility and increasing mortality with advancing age. Such a trait, which impairs survival and fertility, is clearly bad for the individual, raising intriguing questions about why and how it has evolved^{1–4}. Ageing shows a broad phylogenetic distribution but is not universal, as some species show no age-associated increase in mortality or decline in fertility⁵. Thus, ageing cannot be explained simply as the inevitable result of biological wear-and-tear. So, why does it occur?

An early explanation for evolution of ageing was the idea that senescence is programmed in order to limit population size or accelerate the turnover of generations, thereby aiding the adaptation of organisms to changing environments. One essential flaw in this argument is that for most species, other than those like Pacific salmon where death coincides directly with the end of a semelparous (once-only) reproductive cycle, there is scant evidence that senescence contributes significantly to mortality in the wild. Natural mortality, as opposed to that seen in protected populations, is mostly due to extrinsic hazards, such as infection, predation, starvation or cold, and occurs mainly in young individuals. As a rule, wild animals simply do not live long enough to grow old (Fig. 1a). Therefore, natural selection has limited opportunity to exert a direct influence over the process of senescence. Even in species where senescence does make some contribution to mortality in the wild (for example, larger mammals and some long-lived birds), any hypothetical 'accelerated ageing gene' would be disadvantageous to the individual. It is therefore difficult to see how genes for accelerated ageing could be maintained in stable equilibrium, as individuals in whom the genes were inactivated by mutation would enjoy a selection advantage.

The rarity of aged animals in the wild in fact gives the clue to an important principle underlying all of the current evolutionary theories of ageing. As a result of extrinsic mortality, there is a progressive weakening in the force of selection with increasing age⁶. By an age when wild survivorship has declined to very low levels, the force of selection is too weak to oppose the accumulation of germ-line mutations with late-acting deleterious effects⁷ (Fig. 1b). This 'selection shadow' allows a wide range of alleles with late deleterious effects to accumulate over the generations with little or no check. This is the 'mutation accumulation' theory, and because the deleterious alleles are essentially unselected, we might expect considerable heterogeneity in the distribution of such alleles among individuals within the population⁸.

A second theory is that of 'pleiotropy', also sometimes called 'antagonistic pleiotropy'. Williams⁹ suggested that

pleiotropic genes with good early effects would be favoured by selection even if these genes had bad effects at later ages (Fig. 1c). Because the contribution to fitness is a composite of both the size of the effect and the probability of surviving to be affected by it, a small beneficial effect early in life can outweigh a late deleterious effect even if the latter results in senescence and death. This introduces the important idea of a life-history trade-off, which is also a central feature in the third theory, the 'disposable soma' theory^{10,11}, which is based on optimal allocation of metabolic resources between somatic maintenance and reproduction.

Effective somatic maintenance is required only to keep the organism in sound physiological condition for as long as it has a reasonable chance of survival in the wild (Fig. 1d). For example, because more than 90% of wild mice die in their first year¹², any investment in mechanisms for survival beyond this age benefits at most 10% of the population. Nearly all of the mechanisms required to combat intrinsic deterioration (such as DNA repair or antioxidant systems) require metabolic resources. Resources are scarce, as shown by the fact that the main cause of mortality for wild mice is cold, owing to failure to maintain thermogenesis¹³. The disposable-soma theory therefore suggests that the mouse will benefit by investing any spare resource into thermogenesis or reproduction, rather than into better repair capacity, even though this means that damage will eventually accumulate to cause ageing. Although the distinction between the pleiotropy and disposable-soma concepts is sometimes blurred, the latter can be viewed as focusing specifically on mechanisms, particularly the role of somatic maintenance and repair, whereas the former is formulated in terms of a general pattern of gene action and may involve pleiotropic genes of various kinds.

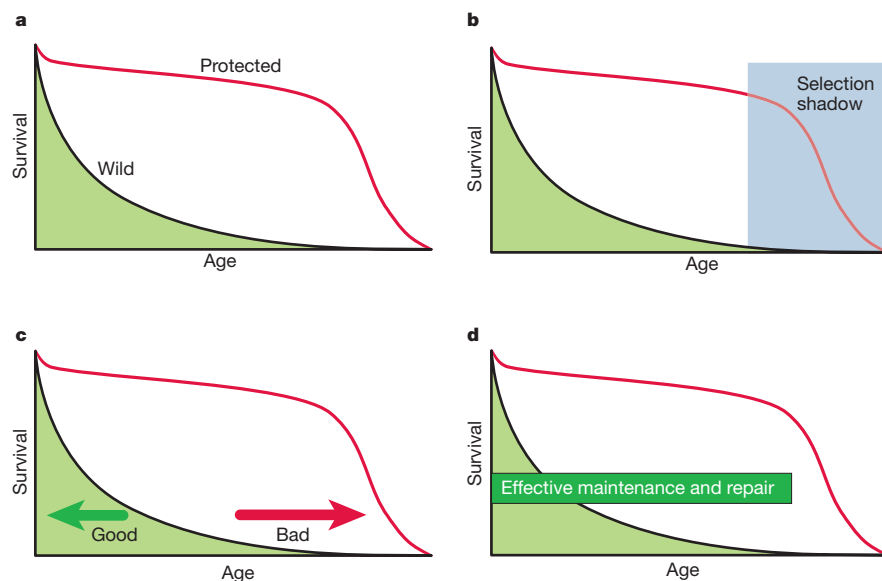
The three theories provide complementary explanations for why ageing occurs. Each also addresses the question: why do species have the life spans they do? The principal determinant in the evolution of longevity is predicted to be the level of extrinsic mortality. If this level is high, life expectancy in the wild is short, the force of selection attenuates fast, deleterious gene effects accumulate at earlier ages, and there is little selection for a high level of somatic maintenance. Consequently, the organism is predicted to be short lived even when studied in a protected environment. Conversely, if the level of extrinsic mortality is low, selection is predicted to postpone deleterious gene effects and to direct greater investment in building and maintaining a durable soma.

Evolutionary theories tested

The evolutionary ageing theories implicitly assume an age-structured population^{6,9}, that is, a population in which

Figure 1 Evolutionary theories of ageing.

a. Extrinsic mortality in wild environments occurs to an extent that senescence-associated mortality is rare, undermining any idea that genes specifically for ageing have evolved. **b.** The 'selection shadow' at older ages may permit an accumulation of late-acting deleterious mutations (mutation-accumulation theory). **c.** Pleiotropic genes that benefit organisms early in life will be favoured by selection even if they have bad effects at later ages (pleiotropy theory). **d.** Selection pressure to invest metabolic resources in somatic maintenance and repair is limited; all that is required is to keep the organism in sound condition for as long as it might survive in the wild (disposable-soma theory).



individuals can be segregated by age, and therefore predict that ageing should not arise in populations where age classes cannot be assigned. At a unicellular level, it is therefore unsurprising that ageing is generally not seen in bacterial populations. Some unicellular organisms show asymmetry of cell division, such as the budding yeast, *Schizosaccharomyces cerevisiae*. Mother yeast cells age in that they show an increasing probability of cell death with successive divisions. In multicellular organisms, the need for age structure also applies and ageing is generally predicted to require a clear separation between germ line and soma^{1,9}. The presence or absence of ageing is sometimes attributed to the presence or absence of sexual reproduction, but this is erroneous. It is the distinction between soma and germ line (a common but not universal correlate of sex) that holds the key. In accord with these predictions, two oligochaete species that reproduce by symmetrical fission were found to show no increase in age-specific mortality, whereas four species (two rotifers, one ostracod and one cladoceran crustacean) which reproduce by asexual egg production all showed highly significant increases¹⁴. *Hydra*, which can reproduce sexually but usually reproduces by asexual budding, and which can regenerate a new individual from almost any part of the organism, lacks clear separation of germ line and soma and shows no obvious signs of intrinsic senescence¹⁵.

An important prediction of the evolutionary theories is that altering the rate of decline in the force of natural selection will lead to the evolution of a concomitantly altered rate of ageing. This has been tested by applying artificial selection on life-history variables or by intra- and interspecies comparisons of populations that are subject to different levels of extrinsic mortality.

Most selection experiments have used the fruitfly *Drosophila melanogaster*. By restricting reproduction to later ages, the intensity of selection on the later portions of the life span was increased. This consistently extended the longevity of the selected populations^{16–19}. Furthermore, a general correlate of delayed senescence has been reduced fecundity in the long-lived flies (Table 1), which supports the idea of a trade-off between fertility and survival, as suggested by the disposable-soma and pleiotropy theories. A similar trade-off was observed in an experiment selecting directly for longevity²⁰. Lastly, when selection operates through the level of extrinsic mortality, populations subjected to low mortality showed increased longevity, longer development times and decreased early fecundity²¹.

The general finding from selection experiments in *Drosophila* is therefore that retarded ageing is associated with depression of fitness components in early life, although there is variation in which fitness

components are affected. Some studies found that body size and development time were increased in the longer-lived lines²¹ whereas other studies fail to observe these effects. Sgró and Partridge²² demonstrated that reduced early fecundity — the most consistently observed feature — was causally involved in the retarded ageing of their long-lived lines; after abolishing reproduction through either irradiation or genetic manipulation they found that the differences in ageing rate between their controls and long-lived selected lines disappeared.

Trade-offs have also been reported in the nematode *Caenorhabditis elegans* where a host of long-lived mutants has been identified. However, although some studies report the same sort of trade-offs between longevity and early-life fitness components as found in *Drosophila*²³, this is not invariably true even for the same genotypes in other laboratories²⁴. A recent study of the long-lived *age-1* mutant in *C. elegans* showed that the relative fitness effects of mutations can be strongly affected by environment²⁵. When mutants were reared together with wild-type individuals under standard culture conditions, neither genotype exhibited a competitive advantage. However, when cultures were alternatively fed and starved — mimicking conditions in nature — the wild type quickly outcompeted the mutant.

Although abundant data support the existence of life-history trade-offs, evidence for the mutation-accumulation theory remains more controversial²⁶. An early *Drosophila* selection experiment indicated a role for mutation accumulation acting alongside trade-offs²⁷, although probably in a minor capacity. Other studies have directly tested the prediction that mutation accumulation should be revealed by an increase in additive genetic variance in mortality rate at later ages. Initial studies in *Drosophila*^{28,29} seemed to support mutation accumulation but later experiments and further analysis have led to doubts about this conclusion^{30,31}.

From the comparative perspective, numerous opportunities exist to test the prediction that in safe environments (those with low extrinsic mortality) ageing will evolve to be retarded, whereas ageing should evolve to be more rapid in hazardous environments. Adaptations that reduce extrinsic mortality (for example, wings, protective shells or large brain) are generally linked with increased longevity (in bats, birds, turtles and humans). Field observations comparing a mainland population of opossums subject to significant predation by mammals, with an island population not subject to mammalian predation, found the predicted slower ageing in the island population³². Among social insect species, those with the most

Table 1 Life-history traits and selection for longevity in *D. melanogaster*

Mode of selection	Traits affected
Delayed reproduction ¹⁶	↑Longevity; ↓early fecundity
Delayed reproduction ¹⁹	↑Longevity; ↓larval viability
Delayed reproduction ¹⁸	↑Longevity; ↓early fecundity
Reduced extrinsic mortality ²¹	↑Longevity; ↓early fecundity; ↑development time
Increased longevity ²⁰	↑Longevity; ↓fecundity

Shown are the life-history traits associated with successful laboratory evolution of increased longevity in *Drosophila*.

protected nests contain reproductive females with by far the longest life spans³³. A comparative analysis of mortality patterns among birds found that the rate of mortality increase with age was directly correlated with the magnitude of presenescent mortality³⁴.

At the molecular and cellular levels, the disposable-soma theory predicts that the proportional effort devoted to cellular maintenance and repair processes will vary directly with longevity. Numerous studies support this idea. For instance, the long-lived rodent species *Peromyscus leucopus* exhibits lower generation of the reactive oxygen species (ROS), which are widely seen as an important contributor to the ageing process, higher cellular concentrations of some antioxidant enzymes, and overall lower levels of protein oxidative damage than the shorter-lived species *Mus musculus*³⁵. A direct relation between species longevity and rate of mitochondrial ROS production in captive mammals has also been found^{36,37}, as has a similar relationship between mammals and similar-sized but much longer-lived birds³⁸. Markers of glycoxidation, the non-enzymatic modification of reducing sugars, are also found to accumulate more slowly in long-lived, as opposed to short-lived, mammals³⁹. DNA repair capacity has been shown to correlate with mammalian life span in numerous comparative studies⁴⁰, as has the level of poly(ADP-ribose) polymerase⁴¹, an enzyme that is important in the maintenance of genomic integrity. The quality of maintenance and repair mechanisms may be revealed by the capacity to cope with external stress. It is notable, therefore, that environmental or genetic manipulations that confer increased longevity on a range of animals also confer increased resistance to environmentally imposed stressors (Table 2). Likewise, comparisons of the functional capacity of cultured cells to withstand a variety of imposed stressors have shown that cells taken from long-lived species have superior stress resistance to that of cells from shorter lived-species^{8,42,43}. All of these studies support the idea that it is the evolved capacity of somatic cells to carry out effective maintenance and repair that governs the time taken for damage to accumulate to levels where it interferes with the organism's viability, and hence regulates longevity.

Specialized life histories

Many organisms live their lives in highly variable environments. In such circumstances we can expect the 'genetic architecture' of the life history, that is, the co-adapted set of traits influencing survival and fecundity, to possess a degree of evolved plasticity that permits a

range of optimal responses suited to different circumstances. In poikilotherms, for instance, environmental temperature often influences longevity. This is generally a straightforward consequence of altering metabolic rate, as indicated by the fact that longevity can be similarly altered by manipulating activity⁴⁴, although the capacity to make such adjustments and retain viability will reflect the range of temperatures to which organisms have been exposed in their evolutionary past. More revealing are cases of plasticity induced by conditions such as lack of food, to which many organisms seem to have evolved a non-reproductive, highly stress-resistant state. Periods of famine often trigger what appear to be metabolic switches that paradoxically extend the normal life span, without sacrificing subsequent reproduction and survival when favourable conditions return⁴⁵.

The most thoroughly investigated instance of this type of life-history plasticity is seen in the nematode *C. elegans*. At 20 °C, wild-type *C. elegans* hermaphrodites live for an average of about 17 days with a maximum of about 25 days. However, under conditions of high larval density and low food availability, larvae develop into the alternative, non-feeding, non-reproducing third instar called the dauer, which can survive for at least 60 days⁴⁶. If conditions improve, dauers moult to adulthood and exhibit normal adult longevity and reproduction. Although dauers exhibit reduced activity and metabolic rate compared with non-dauer larvae^{46,47}, the preservation of full adult longevity after even an extended dauer period indicates that the effect is not entirely due to reduced metabolism. Like longer-lived strains and species of other animals, dauers are also resistant to a variety of environmental stresses including ROS, temperature extremes and ionizing radiation⁴⁸. Increased longevity of adult worms results from mutations in several of the genes involved in the dauer pathway, indicating that some activation of dauer physiology may be involved in these mutants, even in the absence of a triggering shortage of food. Exactly what fraction of the longevity-enhancing effect of dauer formation or long-lived genetic mutants is due to reduced metabolic rate is the subject of current controversy^{23,49}.

It is well known that reduced calorie intake slows ageing in laboratory rodents⁵⁰, an effect that may have parallels with the invertebrate phenomenon. That is, the rodent caloric-restriction response may be a dauer-like adaptive physiological state to 'wait out' periods of food shortage and is generally associated with a partial or complete interruption of fertility. Like many invertebrates, laboratory rodents under caloric restriction show enhanced resistance to a range of stresses⁵¹, but unlike invertebrates, no general reduction in mass-specific metabolic rate is required for the effect. Despite the lack of reduced metabolic rate under caloric restriction, Walford and Spindler⁵² have pointed out a number of biochemical similarities between the caloric-restriction state and that of mammalian hibernation, which can also extend life span. A recent evolutionary model⁵³ has shown that rodents may well have evolved a response to temporary fluctuations in resource availability, in which energy is diverted from reproduction to maintenance functions in periods of food shortage, thereby enhancing survival and retaining reproductive potential for when conditions improve.

Perhaps the most pronounced impact of environment on ageing rate is illustrated by differences between longevity of queens and workers among species of the eusocial insects⁵. Life spans of workers are typically measured in weeks, those of queens in years. Differences in longevity as great as 100-fold have been reported, even though both queens and workers develop from eggs laid by the same mother and fertilized by the same father. The divergence in their relative longevity is mediated by pheromone exposure or the rate and composition of food supplied to the larvae. Although workers are more physically active, queens produce hundreds of thousands of eggs per year, so the difference in ageing rate is unlikely to be due to differences in metabolic expenditure. Indirect evidence indicates that neuroendocrine factors are involved⁵, although underlying these factors must be differences in gene expression. A question as yet

Table 2 Increased stress resistance in long-lived populations

Organism	Population	Nature of stressor				
		ROS	Heat	UV	Trauma	Chemical toxins
<i>C. elegans</i>	Dauer larvae ⁴⁸	↑	↑	↑	?	?
<i>C. elegans</i>	Various mutants ⁴⁹	↑	↑	↑	?	?
<i>D. melanogaster</i>	Artificial selection ^{72,73}	↑	↑	↑	?	?
<i>D. melanogaster</i>	<i>methuselah</i> mutant ⁷⁴	↑	↑	?	?	?
<i>M. musculus</i>	Calorically restricted ⁵¹	↑	↑	?	↑	↑
<i>M. musculus</i>	p66 ^{shc} mutant ⁷⁵	↑*	?	↑†	?	?

*Resistance to apoptosis or growth arrest of cultured embryonic fibroblasts.

†Both *in vitro* and *in vivo* resistance.

unanswered is the extent to which such marked differences in mortality schedules result from purely extrinsic factors (danger to the workers foraging in the environment compared with the relative safety of the queen ensconced in a protected, climate-controlled nest surrounded by self-sacrificing defenders) or from differences in the rate of internal decay. This is an area ripe for research into how differential expression in the same configuration of genes might profoundly affect senescence.

Reproduction and ageing

Most discussion of the evolution of ageing focuses on its effects on mortality, rather than reproduction, in spite of the fact that in terms of an impact on fitness, the form of the reproductive schedule is as important as that of the mortality curve^{54,55}. A good reason for this is that many aspects of reproductive senescence can be explained in the same general terms as physiological senescence. Nevertheless, in addition to the important issues of the trade-offs between survival and fecundity, considered earlier, there are intriguing evolutionary questions about the links between reproduction and ageing, notably the significance of post-reproductive survival (where it occurs) and the effects of damage and selection on the germ line.

The existence of a distinct post-reproductive phase is characteristic of certain semelparous species, in which individuals reproduce only once. On the other hand, many semelparous species undergo extremely rapid senescence on completion of reproduction, often as a direct consequence of the massive physiological changes associated with an explosive reproductive burst⁵. The evolutionary basis of semelparity is well understood⁵⁴, and from the perspective of the evolutionary theories of ageing, semelparity represents an extreme version of the decline in the force of natural selection with age¹. In a semelparous species, the force of natural selection approximates a step function, being uniformly high until reproduction begins, and declining abruptly as reproduction is completed, because the chance of surviving to breed again is effectively zero. This explains the sudden collapse of any pressure to invest in somatic maintenance and repair. Whether or not there is significant post-reproductive survival may be governed chiefly by whether or not the post-reproductive adult contributes actively to the survival chances of the offspring.

A very different example of post-reproductive survival is the human menopause, where fertility in iteroparous human females comes to a relatively abrupt halt at around the age of 45–50 years, when the impact of senescence on most other functions is still small. Although the proximate cause of menopause seems to be oocyte depletion (linked also with neuroendocrine changes), this begs the question why natural selection has not produced a store of oocytes which lasts for longer. One possibility is that during most of humanity's evolutionary history, women rarely survived beyond 45–50 years, so selection simply produced about as many oocytes as would be required. But evidence from hunter–gatherer communities indicates that even though average life expectancy is short, women who avoid the hazards of early life and reach childbearing age have a reasonable chance of surviving to the age of menopause and beyond⁵⁶. This indicates that the menopause may have a deeper evolutionary significance.

Early female reproductive senescence has been reported in other species (for example, chimpanzees, macaques and toothed whales) but is generally less clear-cut, indicating that if the menopause has an evolutionary basis, this may be found in the special circumstances of the human life history. In particular, menopause could be linked with the evolution of human longevity, notably, through the effects of increased brain size and sociality^{9,56–61}. Increased neonatal brain size coupled with the constraint on the birth canal imposed by the mechanics of a bipedal gait has made giving birth unusually difficult for human females. The risks of childbearing, particularly in the absence of modern obstetric care, would increase even more steeply with age if fertility were to persist during the later period of the life span. The problem of a large brain size is also reflected in the fact that

Box 1

Evolutionary genetics of ageing

The evolutionary theories of ageing predict:

1. Specific genes selected to promote ageing are unlikely to exist.
2. Ageing is not programmed but results largely from accumulation of somatic damage, owing to limited investments in maintenance and repair. Longevity is thus regulated by genes controlling levels of activities such as DNA repair and antioxidant defence.
3. In addition, there may be adverse gene actions at older ages arising either from purely deleterious genes that escape the force of natural selection or from pleiotropic genes that trade benefit at an early age against harm at older ages.

It is clear that multiple genes contribute to the ageing phenotype, some being particular to individuals ('private'), others being shared across populations and species ('public')⁸. The principal challenge is to identify how many of each category exist, and which are the most important.

human infants are born unusually early, relative to other species, with respect to the completion of brain growth and development. Infants remain highly dependent for extended periods and, in the ancestral environment, their survival will have been unlikely if their mother died in childbirth. There may thus be a fitness advantage in limiting reproduction to ages when it is comparatively safe, thereby increasing the likelihood of the mother surviving to raise her existing offspring to independence. In addition, post-menopausal women may contribute to the successful rearing of their grandchildren, by providing assistance to their own adult offspring and thereby increasing their inclusive fitness, that is, their overall genetic contribution to future generations. It is likely that a combination of all of these factors is required to explain the human menopause⁶², which may account for the lack of evolutionary support for menopause in other species⁶³.

Human reproductive ageing also highlights the interesting puzzle that although the germ line must, in a fundamental sense, be immortal (as damage cannot be permitted to accumulate across generations without immediate risk of extinction), there is clear evidence that individual germ cells do accumulate faults. Indeed, it would be surprising if the germ line was immune to accumulation of damage, because germ cells are subject to the same kinds of molecular damage as somatic cells. From a statistical perspective it is clear that the germ-cell population does undergo significant ageing. In the case of the human ovary the rate of follicular loss accelerates from around age 35 years, and male fertility begins to decline from around age 30. There is also an increase in the frequency of chromosomal abnormalities in newborn children as a function of maternal and, to a lesser extent, paternal age^{64,65}. Nevertheless, healthy children born to older parents are not prematurely aged, although there is some suggestion that daughters' (but not sons') longevity is adversely affected by advanced paternal age⁶⁵. Thus, either germ cells are endowed with special maintenance and repair systems¹⁰ — the enzyme telomerase being a good example — or selection at the cell or embryo level during gametogenesis, conception and pregnancy serves to screen out most faults. It may be relevant that during each human menstrual cycle about 20 ovarian follicles are triggered to start the process of maturation, although usually only one completes its development and is ovulated. A mechanism of probable significance in the evolution of female germ-line immortality is the stringent bottleneck in the size of the cellular mitochondrial population in early embryogenesis. A healthy complement of mitochondria is essential for subsequent viability of the offspring, and mutations in mitochondrial DNA (mtDNA) tend to accumulate with age⁶⁶. If the mitochondrial population in the oocyte contains a fraction of organelles bearing mtDNA mutations, a bottleneck

coupled with an effective quality screen might select embryos that carry only intact mitochondria⁶⁷.

Implications

Understanding the forces that have sculpted our genetic makeup is likely to provide important insights that can not only guide our investigation of the molecular and cellular basis of ageing, but may also help to identify new routes to positive interventions in the ageing process.

There is a clear prediction that multiple genes influence the ageing process and that these are of several kinds (Box 1). The basis of genetic variation for longevity has begun to be studied in *Drosophila* by estimating quantitative genetic parameters and mapping quantitative trait loci⁶⁸. Gene-expression profiling of ageing rodent tissues is beginning to reveal genes that alter their expression with advancing age or whose expression is altered by interventions, such as caloric restriction, which affect rate of ageing⁶⁹. Unsurprisingly, several of these genes are being identified as those involved in damage/stress-response pathways. Just possibly, such techniques might also reveal late-acting deleterious alleles, as suggested by the mutation-accumulation theory. In general, however, the evolutionary theories caution against interpreting genes whose expression alters in old age as genes that accelerate ageing, and such changes in gene expression are likely to be secondary consequences rather than primary causes of the ageing process.

An important corollary of the prediction (and emerging evidence) that key genes regulating rate of ageing are those that control somatic maintenance and repair, is that at the level of the individual there is considerable scope for the action of stochastic chance⁷⁰. Not only will individual cells within tissues experience different random accumulations of faults, but there may also be important stochastic variations in developmental processes resulting, for example, in different numbers of cells being formed in key organs, such as the hippocampus. Variation in initial cell number and damage rate will in turn affect the time taken before a threshold for dysfunction is crossed during the progressive neurodegeneration that occurs later in life. This means that genetically identical individuals, maintained in uniform environments, may nevertheless exhibit considerable variation in aspects of the senescent phenotype, as has been frequently observed in ageing studies on inbred laboratory organisms⁷⁰. Heterogeneity in the senescent phenotype, arising from intrinsic stochasticity as well as genetic and environmental variations, may also help to explain the intriguing phenomenon that in several species age-specific mortality rate eventually slows its rate of increase and may even decline⁷¹. Heterogeneity explains such an effect if we assume that the frailer individuals die first, leaving a residual population that, as time goes by, represents a dwindling sub-population made up of those individuals that were always the most resilient.

In conjunction with exciting advances in genome analysis, there is clearly much scope for further development and testing of the evolutionary theories of why we age. □

- Kirkwood, T. B. L. & Cremer, T. Cytogerontology since 1881: a reappraisal of August Weismann and a review of modern progress. *Hum. Genet.* **60**, 101–121 (1982).
- Rose, M. R. *Evolutionary Biology of Ageing* (Oxford Univ. Press, New York, 1991).
- Austad, S. N. *Why We Age* (Wiley, New York, 1997).
- Kirkwood, T. B. L. *Time of Our Lives: The Science of Human Ageing* (Oxford Univ. Press, New York, 1999).
- Finch, C. E. *Longevity, Senescence and the Genome* (Chicago Univ. Press, Chicago, 1990).
- Charlesworth, B. *Evolution in Age-Structured Populations* (Cambridge Univ. Press, Cambridge, 1994).
- Medawar, P. B. *An Unsolved Problem of Biology* (Lewis, London, 1952).
- Martin, G. M., Austad, S. N. & Johnson, T. E. Genetic analysis of aging: role of oxidative damage and environmental stresses. *Nature Genet.* **13**, 25–34 (1996).
- Williams, G. C. Pleiotropy, natural selection and the evolution of senescence. *Evolution* **11**, 398–411 (1957).
- Kirkwood, T. B. L. Evolution of ageing. *Nature* **270**, 301–304 (1977).
- Kirkwood, T. B. L. Human senescence. *BioEssays* **18**, 1009–1016 (1996).
- Phelan, J. P. & Austad, S. N. Natural selection, dietary restriction and extended longevity. *Growth Dev. Aging* **53**, 4–6 (1989).
- Berry, R. J. & Bronson, F. H. Life history and bioeconomy of the house mouse. *Biol. Rev.* **67**, 519–550 (1992).

- Bell, G. Evolutionary and nonevolutionary theories of senescence. *Am. Nat.* **124**, 600–603 (1984).
- Martinez, D. E. Mortality patterns suggest lack of senescence in hydra. *Exp. Gerontol.* **33**, 217–225 (1997).
- Rose, M. R. & Charlesworth, B. A test of evolutionary theories of senescence. *Nature* **287**, 141–142 (1980).
- Luckinbill, L. S., Arking, R., Clare, M. J., Cirocco, W. C. & Buck, S. A. Selection of delayed senescence in *Drosophila melanogaster*. *Evolution* **38**, 996–1003 (1984).
- Partridge, L., Prowse, N. & Pignatelli, P. Another set of responses and correlated responses to selection on age of reproduction in *Drosophila melanogaster*. *Proc. R. Soc. Lond. B* **266**, 255–261 (1999).
- Buck, S., Vetrano, J., Force, A. G. & Arking, R. Extended longevity in *Drosophila* is consistently associated with a decrease in larval viability. *J. Gerontol. A* **55**, B292–B301 (2000).
- Zwaan, B. J., Bijlsma, R. & Hoekstra, R. F. Direct selection on lifespan in *Drosophila melanogaster*. *Evolution* **49**, 649–659 (1995).
- Stearns, S. C., Ackermann, M., Doebeli, M. & Kaiser, M. Experimental evolution of aging, growth, and reproduction in fruitflies. *Proc. Natl Acad. Sci. USA* **97**, 3309–3313 (2000).
- Sgró, C. M. & Partridge, L. A delayed wave of death from reproduction in *Drosophila*. *Science* **286**, 2521–2524 (1999).
- Van Voorheis, W. A. & Ward, S. Genetic and environmental conditions that increase longevity in *Caenorhabditis elegans* decrease metabolic rate. *Proc. Natl Acad. Sci. USA* **95**, 11399–11403 (1999).
- Johnson, T. E. & Hutchinson, E. W. Absence of strong heterosis for life span and other life history traits in *Caenorhabditis elegans*. *Genetics* **134**, 465–474 (1993).
- Walker, D. W., McColl, G., Jenkins, N. L., Harris, J. & Lithgow, G. J. Evolution of lifespan in *C. elegans*. *Nature* **405**, 296–297 (2000).
- Stearns, S. C. & Partridge, L. in *Handbook of the Biology of Aging* (eds Masoro, E. J. & Austad, S. N.) (Academic, San Diego, in the press).
- Service, P. M., Hutchinson, E. W. & Rose, M. R. Multiple genetic mechanisms for the evolution of senescence in *Drosophila melanogaster*. *Evolution* **42**, 708–716 (1988).
- Hughes, K. A. & Charlesworth, B. A genetic analysis of senescence in *Drosophila*. *Nature* **367**, 64–66 (1994).
- Charlesworth, B. & Hughes, K. A. Age-specific inbreeding depression and components of genetic variance in relation to the evolution of senescence. *Proc. Natl Acad. Sci. USA* **93**, 6140–6145 (1996).
- Promislow, D. E. L., Tatar, M., Khazaeli, A. A. & Curtsinger, J. W. Age-specific patterns of genetic variation in *Drosophila melanogaster*. I. Mortality. *Genetics* **143**, 839–848 (1995).
- Shaw, F. H., Promislow, D. E. L., Tatar, M., Hughes, K. A. & Geyes, C. J. Toward reconciling inferences concerning genetic variation in *Drosophila melanogaster*. *Genetics* **152**, 553–566 (1999).
- Austad, S. N. Retarded senescence in an insular population of opossums. *J. Zool.* **229**, 695–708 (1993).
- Keller, L. & Genoud, M. Extraordinary lifespans in ants: a test of evolutionary theories of aging. *Nature* **389**, 958–960 (1997).
- Ricklefs, R. E. Evolutionary theories of aging: confirmation of a fundamental prediction, with implications for the genetic basis and evolution of life span. *Am. Nat.* **152**, 24–44 (1998).
- Sohal, R. S., Ku, H.-H. & Agarwal, S. Biochemical correlates of longevity in two closely-related rodent species. *Biochem. Biophys. Res. Comm.* **196**, 7–11 (1993).
- Ku, H.-H., Brunk, U. T. & Sohal, R. S. Relationship between mitochondrial superoxide and hydrogen-peroxide production and longevity of mammalian-species. *Free Radical Biol. Med.* **15**, 621–627 (1993).
- Barja, G. & Herrero, A. Oxidative damage to mitochondrial DNA is inversely related to maximum life span in the heart and brain of mammals. *FASEB J.* **14**, 312–318 (2000).
- Herrero, A. & Barja, G. 8-oxo-deoxyguanosine levels in heart and brain mitochondrial and nuclear DNA of two mammals and three birds in relation to their different rates of aging. *Aging Clin. Exp. Res.* **11**, 294–300 (1999).
- Sell, D. R. et al. Longevity and the genetic determination of collagen glycoxidation kinetics in mammalian senescence. *Proc. Natl Acad. Sci. USA* **93**, 485–490 (1996).
- Kirkwood, T. B. L. DNA, mutations and aging. *Mut. Res.* **219**, 1–7 (1989).
- Grube, K. & Bürkle, A. Poly(ADP-ribose) polymerase activity in mononuclear leukocytes of 13 mammalian species correlates with species-specific life span. *Proc. Natl Acad. Sci. USA* **89**, 11759–11763 (1992).
- Ogburn, C. E. et al. Cultured renal epithelial cells from birds and mice: enhanced resistance of avian cells to oxidative stress and DNA damage. *J. Gerontol.* **53**, B287–B292 (1998).
- Kapahi, P., Boulton, M. E. & Kirkwood, T. B. L. Positive correlation between mammalian life span and cellular resistance to stress. *Free Radical Biol. Med.* **26**, 495–500 (1999).
- Sohal, R. S. & Donato, H. Effects of experimentally altered life span on the accumulation of fluorescent age pigment in the housefly, *Musca domestica*. *Exp. Gerontol.* **13**, 335–341 (1978).
- Austad, S. N. Life extension by dietary restriction in the bowl and doily spider, *Frontinella pyramitella*. *Exp. Gerontol.* **24**, 83–92 (1989).
- Klass, M. & Hirsh, D. Non-ageing developmental variant of *Caenorhabditis elegans*. *Nature* **260**, 523–525 (1976).
- Anderson, G. L. Responses of dauer larvae of *Caenorhabditis elegans* (Nematoda: Rhabditidae) to thermal stress and oxygen deprivation. *Can. J. Zool.* **56**, 1786–1791 (1978).
- Lithgow, G. J. in *Handbook of the Biology of Aging* (eds Schneider, E. L. & Rowe, J. W.) 55–73 (Academic, San Diego, 1996).
- Vanfleteren, S. L. & De Vreese, A. Rate of aerobic metabolism and superoxide production rate potential in the nematode *Caenorhabditis elegans*. *J. Exp. Zool.* **274**, 93–100 (1996).
- Weindruch, R. & Walford, R. L. *The Retardation of Aging and Disease by Dietary Restriction* (Thomas, Springfield, IL, 1988).
- Masoro, E. J. Influence of caloric intake on aging and on the response to stressors. *J. Toxicol. Environ. Health B Crit. Rev.* **1**, 243–257 (1998).
- Walford, R. L. & Spindler, S. R. The response to calorie restriction in mammals shows features also common to hibernation: a cross-adaptation hypothesis. *J. Gerontol.* **52**, B179–B183 (1997).
- Shanley, D. P. & Kirkwood, T. B. L. Calorie restriction and aging: a life history analysis. *Evolution* **54**, 740–750 (2000).
- Stearns, S. C. *The Evolution of Life Histories* (Oxford Univ. Press, New York, 1992).
- Partridge, L. & Barton, N. H. On measuring the rate of ageing. *Proc. R. Soc. Lond. B* **263**, 1365–1371 (1996).
- Hill, K. & Hurtado, A. M. *Ache Life History: the Ecology and Demography of a Foraging People* (de Gruyter, New York, 1996).

57. Hamilton, W. D. The moulding of senescence by natural selection. *J. Theor. Biol.* **12**, 12–14 (1966).
58. Kirkwood, T. B. L. & Holliday, R. in *The Biology of Human Ageing* (eds Collins, A. J. & Bittles, A. H.) 1–16 (Cambridge Univ. Press, Cambridge, 1986).
59. Austad, S. N. Menopause: an evolutionary perspective. *Exp. Gerontol.* **29**, 255–263 (1994).
60. Peccei, J. S. The origin and evolution of menopause: the altriciality-lifespan hypothesis. *Ethol. Sociobiol.* **16**, 425–449 (1995).
61. Hawkes, K., O'Connell, J. F., Jones, N. G. B., Alvarez, H. & Charnov, E. L. Grandmothering, menopause, and the evolution of human life histories. *Proc. Natl Acad. Sci. USA* **95**, 1336–1339 (1998).
62. Shanley, D. P. & Kirkwood, T. B. L. Evolution of the human menopause. *BioEssays* (in the press).
63. Packer, C., Tatar, M. & Collins, A. Reproductive cessation in female mammals. *Nature* **392**, 807–822 (1998).
64. Risch, N., Reich, E. W., Wishnick, M. M. & McCarthy, J. G. Spontaneous mutation and parental age in humans. *Am. J. Hum. Genet.* **41**, 218–248 (1987).
65. Gavrilov, L. A. & Gavrilova, N. S. Parental age at conception and offspring longevity. *Rev. Clin. Gerontol.* **7**, 5–12 (1997).
66. Wallace, D. C. Mitochondrial diseases in man and mouse. *Science* **283**, 1482–1488 (1999).
67. Cummins, J. M. in *Female Reproductive Aging* (eds Te Velde, E. R., Pearson, P. L. & Broekmans, F. J.) 207–224 (Parthenon, Carnforth, UK, 2000).
68. Vieira, C. et al. Genotype-environment interaction for quantitative trait loci affecting life span in *Drosophila melanogaster*. *Genetics* **154**, 213–227 (2000).
69. Lee, C. K., Klopp, R. G., Weindruch, R. & Prolla, T. A. Gene expression profile of aging and its retardation by caloric restriction. *Science* **285**, 1390–1393 (1999).
70. Finch, C. E. & Kirkwood, T. B. L. *Chance, Development and Aging* (Oxford Univ. Press, New York, 2000).
71. Vaupel, J. W. et al. Biodemographic trajectories of longevity. *Science* **280**, 855–860 (1998).
72. Service, P. M., Hutchinson, E. W., Mackinley, M. D. & Rose, M. R. Resistance to environmental-stress in *Drosophila melanogaster* selected for postponed senescence. *Physiol. Zool.* **58**, 380–389 (1985).
73. Tatar, M. Evolution of senescence: longevity and the expression of heat shock proteins. *Am. Zool.* **39**, 920–927 (1999).
74. Lin, Y. J., Seroude, L. & Benzer, S. Extended life-span and stress resistance in the *Drosophila* mutant *methuselah*. *Science* **282**, 943–946 (1998).
75. Migliaccio, E. et al. The p66^{bac} adaptor protein controls oxidative stress response and life span in mammals. *Nature* **402**, 309–313 (1999).

Oxidants, oxidative stress and the biology of ageing

Toren Finkel* & Nikki J. Holbrook†

*Laboratory of Molecular Biology, National Heart, Lung, and Blood Institute/National Institutes of Health, Building 10/6N-240, 10 Center Drive, Bethesda, Maryland 20892-1622, USA (e-mail: finkelt@nih.gov)

†Laboratory of Biological Chemistry, National Institute on Aging/National Institutes of Health, 5600 Nathan Shock Drive, Baltimore, Maryland 21224, USA (e-mail: nikki-holbrook@nih.gov)

Living in an oxygenated environment has required the evolution of effective cellular strategies to detect and detoxify metabolites of molecular oxygen known as reactive oxygen species. Here we review evidence that the appropriate and inappropriate production of oxidants, together with the ability of organisms to respond to oxidative stress, is intricately connected to ageing and life span.

Nearly a century ago it was noted that animals with higher metabolic rates often have shorter life spans. These observations led to the formulation of 'the rate-of-living hypothesis', which states that the metabolic rate of a species ultimately determines its life expectancy. Initially, the mechanistic link between metabolism and ageing was unknown. In the mid-1950s, Denham Harman articulated a 'free-radical theory' of ageing, speculating that endogenous oxygen radicals were generated in cells and resulted in a pattern of cumulative damage¹. Although the concept of endogenous oxidants was at first controversial, the identification a decade later of superoxide dismutase (SOD)², an enzyme whose sole function seemed to be the removal of superoxide anions, provided mechanistic support for Harman's hypothesis. Given that the mitochondria produce most of the energy in the cell, and correspondingly consume the bulk of intracellular oxygen, the free-radical theory of ageing is now often thought of synonymously with the rate-of-living hypothesis; the higher the metabolic rate of an organism, the greater the production of reactive oxygen species (ROS) and hence the shorter the life span. However, in some species the strict correlation between metabolic rate and life span is not maintained. This is particularly true for birds and primates, which tend to live longer than would be predicted by their metabolic rates. Careful analysis of oxidant production demonstrated that at a given metabolic rate, mitochondria from these species tend to produce fewer ROS³. This indicates that ROS production rather than metabolic rate provides the strongest correlation with overall longevity.

The free-radical theory of ageing originally implied that the targets of ROS were random, indiscriminate and cumulative. However, although oxidants may certainly function stochastically, accumulating evidence implicates ROS as specific signalling molecules under both physiological and pathophysiological conditions. This more complex view of the importance of oxidants in biological processes is depicted in Fig. 1. As indicated, the generation of ROS, within certain boundaries, is essential to maintain homeostasis. For example, ROS generation by phagocytic cells constitutes an essential host defence mechanism necessary to combat infection. Likewise, cytosolic ROS produced in response to stimulation by growth factors are involved in regulating the proliferative response⁴. Under certain situations of metabolic stress, even mitochondrial-derived oxidants

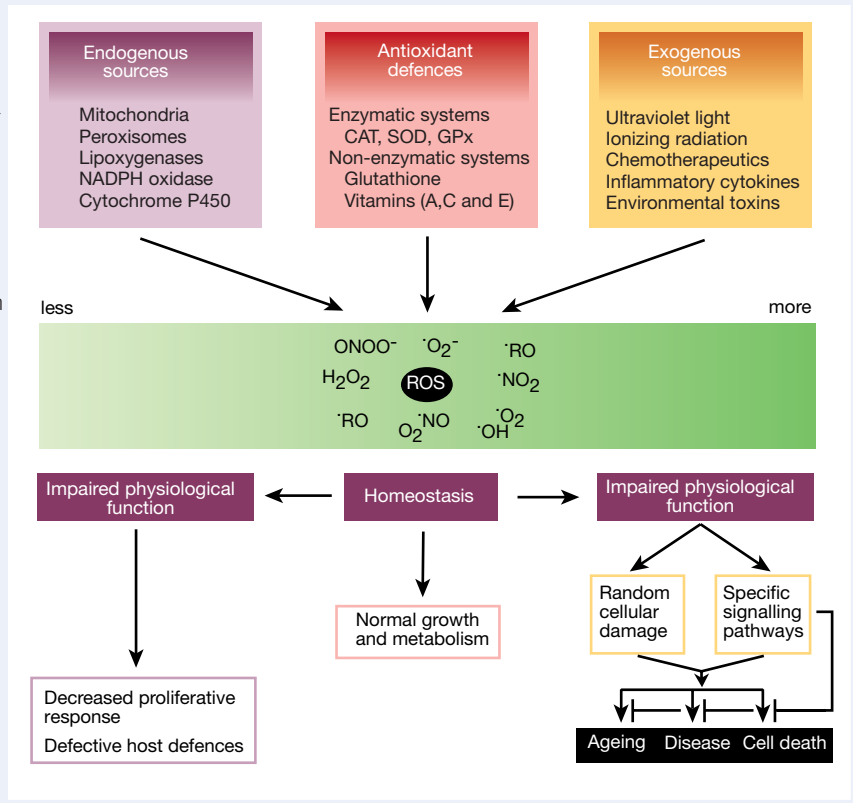
seem to function as signalling molecules^{5,6}. Regardless of how or where they are generated, a rise in intracellular oxidant levels has two potentially important effects: damage to various cell components and triggering of the activation of specific signalling pathways. Both of these effects can influence numerous cellular processes linked to ageing and the development of age-related diseases. In this review, we address how ROS are generated, how the cell responds to oxidative stress and how these responses change with age. In addition, we describe the mounting genetic evidence that links oxidants and oxidative stress responsiveness to ageing and discuss the challenges associated with the potential development of anti-ageing therapies.

Oxidant production and antioxidant defences

For the purpose of this discussion, ROS encompass a variety of diverse chemical species including superoxide anions, hydroxyl radicals and hydrogen peroxide. Some of these species, such as superoxide or hydroxyl radicals, are extremely unstable, whereas others, like hydrogen peroxide, are freely diffusible and relatively long-lived. These various radical species can either be generated exogenously or produced intracellularly from several different sources. Cytosolic enzyme systems contributing to oxidative stress include, among others, the expanding family of NADPH oxidases, a superoxide-generating system that was first described in the neutrophil. The newly described non-phagocytic NADPH oxidases also generate superoxide anions, but depending on the specific NADPH oxidase expressed, can either trigger cellular transformation or replicative senescence^{7,8}. The observation that different members of the NADPH oxidase family can have such widely differing biological outcomes reinforces the complexity in determining the cellular response to oxidants. Contributing factors may include the cell type, the absolute level and duration of oxidant production, the species of ROS generated, and the specific intracellular site of ROS production. Nonetheless, the NADPH oxidase family of enzymes, like the widely characterized nitric oxide synthase (NOS) family, illustrates the apparent purposeful and deliberate use of oxidant generation in normal cellular signalling and homeostasis.

Most estimates suggest that the majority of intracellular ROS production is derived from the mitochondria. The production of mitochondrial superoxide radicals occurs primarily at two discrete points in the electron transport chain, namely at complex I (NADH dehydrogenase) and at

Figure 1 The sources and cellular responses to reactive oxygen species (ROS). Oxidants are generated as a result of normal intracellular metabolism in mitochondria and peroxisomes, as well as from a variety of cytosolic enzyme systems. In addition, a number of external agents can trigger ROS production. A sophisticated enzymatic and non-enzymatic antioxidant defence system including catalase (CAT), superoxide dismutase (SOD) and glutathione peroxidase (GPx) counteracts and regulates overall ROS levels to maintain physiological homeostasis. Lowering ROS levels below the homeostatic set point may interrupt the physiological role of oxidants in cellular proliferation and host defence. Similarly, increased ROS may also be detrimental and lead to cell death or to an acceleration in ageing and age-related diseases. Traditionally, the impairment caused by increased ROS is thought to result from random damage to proteins, lipids and DNA. In addition to these effects, a rise in ROS levels may also constitute a stress signal that activates specific redox-sensitive signalling pathways. Once activated, these diverse signalling pathways may have either damaging or potentially protective functions.



complex III (ubiquinone–cytochrome *c* reductase). Under normal metabolic conditions, complex III is the main site of ROS production⁹. With respect to human ageing, the Achilles' heel of this elegant system lies in the formation of the free radical semiquinone anion species ($\cdot Q^-$) that occurs as an intermediate in the regeneration of coenzyme Q (Fig. 2). Once formed, $\cdot Q^-$ can readily and non-enzymatically transfer electrons to molecular oxygen with the subsequent generation of a superoxide radical. The generation of ROS therefore becomes predominantly a function of metabolic rate and, as such, the rate of living can be indirectly translated to a corresponding rate of oxidative stress. In addition to generating oxidants, metabolism can produce a host of other by-products including glyoxal and methylglyoxal, both of which can contribute to advanced glycation end-product (AGE) formation that, in turn, seems to contribute to the ageing phenotype¹⁰.

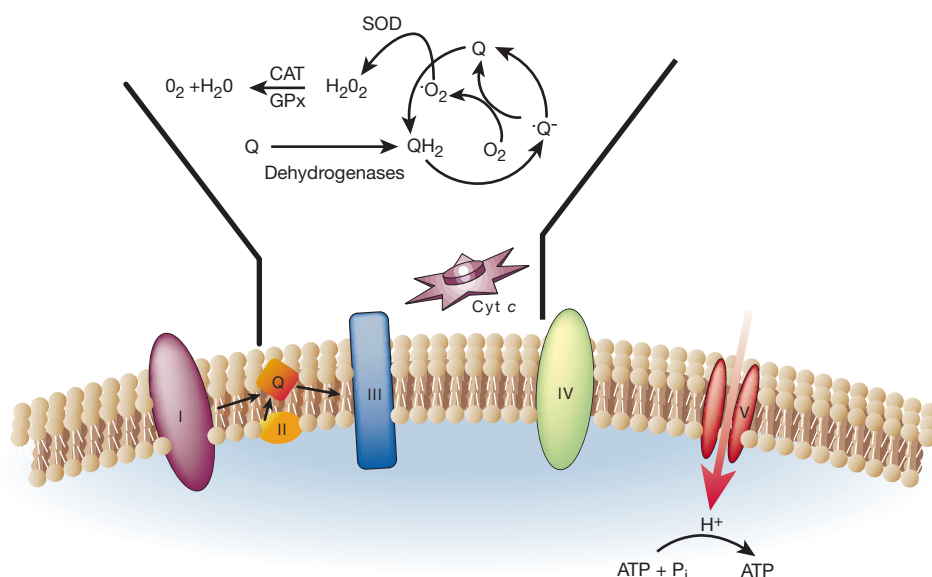
Evidence indicates that, *in vitro*, mitochondria convert 1–2% of the oxygen molecules consumed into superoxide anions¹¹. Given that these initial estimates were made on isolated mitochondria in the presence of high, non-physiological concentrations of oxygen, the *in vivo* rate of mitochondrial superoxide production is undoubtedly considerably less. Whatever the absolute amount of mitochondrial ROS, given their potentially harmful effects, it is likely that numerous protective mechanisms have evolved to limit oxidant production and release. A further understanding of such regulatory mechanisms may reveal potential targets for therapeutic intervention. One postulated mechanism to reduce mitochondrial oxidant production is to increase the rate of metabolic uncoupling¹². When oxygen consumption is uncoupled from ATP generation, heat is produced. This thermogenesis is mediated by an expanding family of uncoupling proteins (UCP-1, -2 and -3). However, the consumption of oxygen without ATP production would also reduce the level of free molecular oxygen potentially available for superoxide anion formation (Fig. 2). Consistent with the hypothesis that metabolic uncoupling might regulate ROS release is the recent evidence indicating that an increase in uncoupling reduces mitochondrial ROS release⁵, whereas levels of mitochondrial oxidants rise in mice with a targeted deletion of UCP-3 (ref. 13).

The burden of ROS production is largely counteracted by an intricate antioxidant defence system that includes the enzymatic scavengers SOD, catalase and glutathione peroxidase. SOD speeds the conversion of superoxide to hydrogen peroxide, whereas catalase and glutathione peroxidase convert hydrogen peroxide to water. In addition to these well characterized antioxidant enzymes, at least five members of a new family of peroxide scavengers termed peroxiredoxins have recently been isolated¹⁴. A variety of other non-enzymatic, low molecular mass molecules are important in scavenging ROS. These include ascorbate, pyruvate, flavonoids, carotenoids and perhaps most importantly, glutathione, which is present in millimolar concentrations within cells.

The balance between ROS production and antioxidant defences determines the degree of oxidative stress. Consequences of this stress include modification to cellular proteins, lipids and DNA. The most widely studied oxidative stress-induced modification to proteins is the formation of carbonyl derivatives¹⁵. Carbonyl formation can occur through a variety of mechanisms including direct oxidation of certain amino-acid side chains and oxidation-induced peptide cleavage. Although all organs and all proteins can potentially be modified by oxidative stress, certain tissues and specific protein targets may be especially sensitive^{16,17}. A recent report indicated that protein misfolding, independent of the cellular redox state, increases protein carbonylation¹⁸. As such, the notion that the rate of carbonyl formation is always directly proportional to the degree of oxidative stress may need to be re-examined.

Several studies have shown that ageing cells and organisms accumulate increased levels of oxidant-damaged nuclear DNA¹⁹. Perhaps because of its proximity to the main source of oxidant generation, or because of a limited DNA repair system, mitochondrial DNA is generally considered to be even more sensitive than nuclear DNA to oxidative damage. Two recent studies provided direct proof that oxidative stress can induce mitochondrial DNA damage. In these studies, oxidative stress was genetically engineered by targeted deletions in either Mn-SOD or the adenine nuclear transporter. These knockout mice had a respective defect in mitochondrial

Figure 2 Complex III is the major source of mitochondrial ROS production. Electrons from complex I or II dehydrogenases are transferred to coenzyme Q (Q), also called ubiquinone. The resulting reduced form (QH₂) of coenzyme Q subsequently undergoes two sequential one-electron reductions (the Q cycle) using oxidized and reduced forms of cytochrome *b* and cytochrome *c* (Cyt *c*). The unstable intermediate in the Q cycle (·Q·) can lead to superoxide formation by transferring electrons directly to molecular oxygen. Once generated, the superoxide can be enzymatically dismutated by SOD to form hydrogen peroxide that in turn is metabolized by enzymes such as catalase (CAT) and glutathione reductase (GPx) regenerating water and molecular oxygen. The generation of superoxide is non-enzymatic and hence the higher the rate of metabolism, the greater the production of ROS.



superoxide scavenging capacity or overall aerobic metabolism. Subsequent analysis of these animals showed significant increase in the level of mitochondrial DNA rearrangements^{20,21}. Increasing damage to mitochondrial DNA inevitably leads to compromised mitochondrial function and integrity. Damaged mitochondria are thought to release more ROS and set in motion a vicious cycle of increasing DNA damage leading to increased ROS production that in turn leads to more DNA damage.

Cellular senescence and oxidative stress

After a finite number of divisions, primary cell cultures enter a state of replicative senescence in which they are growth-arrested and refractory to further mitogenic stimulation. Although the relevance of *in vitro* senescence to organismal ageing remains controversial, several studies indicate that oxidants are important in the development of the senescent phenotype. Early studies with human diploid fibroblasts revealed that cells grown in low oxygen tension exhibit a prolonged life span²². In contrast, cells grown in the presence of high oxygen concentrations have a reduced life span and show an accelerated rate of telomere shortening per population doubling²³. Similarly, treatment of cultures of primary fibroblasts with moderate, non-lethal doses of exogenous hydrogen peroxide activates a rapid, senescence-like growth arrest²⁴.

The role of oxidants in cellular senescence was further underscored by recent observations that overexpression of an activated Ras gene can also induce a senescence-like state in human diploid fibroblasts²⁵. Subsequent analysis demonstrated that expression of activated Ras in diploid fibroblasts resulted in an increase in oxidant levels²⁶. In addition, although expression of activated Ras in human diploid fibroblasts produced growth arrest, this arrest could be reversed either by reducing ambient oxygen or by treatment with a cell-permeable antioxidant. As such, these results raise the possibility that a moderate, sustained rise in oxidants may function as a common trigger for activation of the senescence programme.

Oxidants in cellular signalling

Although the preceding discussions have focused mainly on the endogenous generation of ROS as a consequence of metabolic activities, many environmental stimuli including cytokines, ultraviolet (UV) radiation, chemotherapeutic agents, hyperthermia and even growth factors generate high levels of ROS that can perturb the normal redox balance and shift cells into a state of oxidative stress.

When the stress is severe, survival is dependent on the ability of the cell to adapt to or resist the stress, and to repair or replace the damaged molecules. Alternatively, cells may respond to the insult by undergoing apoptosis, a process whereby severely damaged cells are removed from the multicellular host, and which, within limits, preserves the organism. A number of stress response mechanisms have evolved to help the cell and organism adapt to acute stress, and acting in either a cooperative or antagonistic fashion they serve to coordinate the acute cellular stress response and ultimately determine the outcome. Many of these pathways have been faithfully preserved throughout evolution. Thus, both the stresses and the response mechanisms represent players in an ancient battle. As will be discussed later, many mutations that prolong life seem to provide a global increase in stress resistance. Therefore, a more complete understanding of the cellular response to stress should provide significant insight into ageing.

Among the main stress signalling pathways and/or central mediators activated in response to oxidant injury are the extracellular signal-regulated kinase (ERK), c-Jun amino-terminal kinase (JNK) and p38 mitogen-activated protein kinase (MAPK) signalling cascades, the phosphoinositide 3-kinase (PI(3)K)/Akt pathway, the nuclear factor (NF)-κB signalling system, p53 activation, and the heat shock response (Fig. 3). Activation of these pathways is not unique to oxidative stress, as they are known to have central roles in regulating cellular responses to other stresses as well as regulating normal growth and metabolism. Indeed, in some situations the response to oxidants may involve overstimulation of normal ROS-regulated signalling pathways. In general, the heat shock response, ERK, PI(3)K/Akt and NF-κB signalling pathways exert a pro-survival influence during oxidant injury, whereas activation of p53, JNK and p38 are more commonly linked to apoptosis. However, numerous exceptions to these generalities can be found.

The initiating events leading to activation of pathways in response to oxidants are incompletely understood. In the case of p53 activation, oxidative stress may be sensed as a consequence of the DNA damage it causes. However, in some cells, elevated p53 expression results in increased oxidative stress^{27,28}, suggesting that an important consequence of oxidant-induced p53 activation is a further increase in the level of oxidative stress. This positive feedback loop may be important in committing to an apoptotic response.

Oxidants seem to activate the ERK and the PI(3)K/Akt pathways largely through stimulation of growth-factor receptors, mimicking

Table 1 Mutations revealing links between longevity and stress resistance in multicellular organisms

Species/mutations	Gene description	Life span†	Phenotypic influence*	Stress resistance
C. elegans				
<i>age-1</i>	Human PI(3)K homologue	65% increase		Enhanced (UV, paraquat, heat)
<i>daf-2</i>	Human insulin-receptor homologue	100% increase		Enhanced (UV, paraquat, heat)
<i>daf-16</i>	Forkhead transcription factor	Suppresses longevity conferred by <i>age-1</i> and <i>daf-2</i> mutations		Suppresses stress resistance of <i>age-1</i> and <i>daf-2</i> mutants
<i>clk-1</i>	Homologue of yeast gene associated with coenzyme Q biosynthesis	40% increase		Enhanced increased (UVC)
<i>spe-10</i>	Unknown (sperm defective)	40% increase		Enhanced (UV, paraquat, but not heat)
<i>spe-26</i>	Unknown (sperm defective)	65% increase		Enhanced (UV, paraquat, heat)
<i>old-1</i>	Putative receptor tyrosine kinase	65% increase		Enhanced (UV, heat)
<i>ctl-1</i>	Cytosolic catalase	25% decrease; suppresses longevity conferred by <i>daf-2</i> , <i>age-1</i> and <i>clk-1</i>		Not determined
<i>mev-1</i>	Cytochrome <i>b</i> subunit of succinate dehydrogenase	37% decrease		Hypersensitive to oxygen
Drosophila				
<i>mth</i>	Putative G-protein-coupled receptor	35% increase		Enhanced (UV, paraquat, heat)
Mouse				
<i>shc</i> ⁶⁶	Cytoplasmic signal-transduction adaptor protein	30% increase		Enhanced (UV, H ₂ O ₂)

*See ref. 83 for original references describing the phenotypes of the *daf-2*, *daf-16*, *clk-1* and *spe-26* mutants. References for other mutants are as follows: *age-1* (refs 49, 50, 83); *spe-10* (ref. 84); *old-1* (ref. 85); *ctl-1* (ref. 56); *mev-1* (ref. 55); *mth* (ref. 60); and *shc*⁶⁶ (ref. 68).

†Numbers reflect changes in mean life span of the mutants relative to wild-type animals.

the actions of natural ligands. Many growth-factor receptors have been shown to undergo enhanced phosphorylation in response to direct treatment with oxidants, and agents or conditions that prevent receptor phosphorylation likewise inhibit the activation of ERK and Akt by oxidants^{29–31}. One mechanism proposed to explain this effect is oxidant-mediated inactivation of critical phosphatases necessary for dephosphorylation (turning off) of the growth-factor receptors³¹. Support for such a mechanism has come from the finding that hydrogen peroxide, either derived exogenously or produced endogenously after growth-factor stimulation, can reversibly inactivate protein-tyrosine phosphatase 1B in cells³². The activation of growth-factor-receptor signalling pathways by oxidants is consistent with the demonstration that low concentrations of exogenous hydrogen peroxide are mitogenic³³.

Oxidative stress might induce activation of the JNK and p38 kinase pathways by an additional mechanism. The redox regulatory protein thioredoxin (Trx) has been shown to bind to apoptosis signal-regulating kinase (ASK1), an upstream activator of both JNK and p38, and under normal conditions inhibit its activity³⁴. However, oxidative stress causes dissociation of the Trx–ASK1 complex and subsequent activation of the downstream JNK and p38 kinase³⁴. Similarly, biochemical evidence indicates that under non-stressed conditions glutathione *S*-transferase binds to JNK to inhibit its activation, but that this interaction is also disrupted by oxidative stress³⁵. These results show an intimate coupling between alterations in the intracellular redox state and the activity of downstream stress-activated pathways. The observation that multiple pathways are sensitive to a rise in ROS levels indicates that these pathways may have evolved, in part, to allow organisms to survive within an aerobic environment. In addition, it suggests that a rise in ROS might represent a common, if not universal, signal of cellular stress.

A common effect of the activation of these pathways is a change in pattern of gene expression mediated largely through modulation of the activities of transcription factors. Accordingly, a large number of oxidative stress-responsive transcription factors and genes have been identified³⁶ and some of these have been implicated in influencing ageing processes. The effect of ROS on expression and activity of transcription factors is complex and occurs at multiple levels, often in a seemingly antagonistic or paradoxical fashion. For example, although ROS generally cause an increase in AP-1 levels and increased nuclear translocation of NF- κ B, oxidant stress can at the same time reduce the transcriptional activity of these molecules

through the direct oxidation of critical cysteine residues contained within the DNA-binding domain³⁶.

Analysis of the integrated cellular response to oxidative stress is an area that is particularly suited for analysis using a genomic or proteomic approach. Initial characterization using the latter technique in peroxide-challenged yeast revealed over a 100 proteins whose levels changed after oxidative stress³⁷. As might be expected, included among these were proteins involved in scavenging ROS, as well as heat shock proteins and chaperones. Nearly a quarter of the proteins identified were involved in carbohydrate metabolism. In general, oxidative stress repressed a number of proteins involved in glycolysis and the tricarboxylic acid cycle. This was interpreted as an attempt to restore levels of reducing equivalents such as NADPH at the expense of ATP generation. If these results are confirmed in higher eukaryotes it would imply that ROS may not only be by-products of metabolism but also be regulators of metabolic rates. Such a concept further blurs the distinction between the rate-of-living hypothesis and the free-radical theory of ageing and raises the possibility that the rate of living may determine the level of oxidant generation and that oxidant generation may in turn modulate the rate of living.

Several of the pathways activated by acute oxidative stress show diminished activity as a function of ageing. In a variety of model systems and stress paradigms including oxidative stress, the magnitude of induction of heat-shock proteins, and Hsp70 in particular, is attenuated with age³⁸. The heat-shock protein family encompasses many chaperones involved in regulating the folding, transport and degradation of other cellular proteins. The relationship between ageing and a decline in the robustness of this stress response is unclear, but evidence indicates that elevating levels of Hsp70 enhances the survival of stressed cells and/or animals, whereas inhibiting this stress response reduces survival³⁹.

In contrast to the age-related attenuation in heat-shock protein induction in response to acute stress, two recent studies using complementary DNA microarray analysis to examine global age-associated changes in gene expression in mouse tissues provided evidence that basal expression (that seen in the absence of overt stress) of certain heat-shock proteins actually increases with ageing^{40,41}. This elevated expression (which involves members other than Hsp70) was interpreted to occur as a response to the age-associated accumulation of oxidatively damaged proteins. Age-associated elevations in expression of heat-shock proteins have also been observed in *Drosophila*⁴² and *Caenorhabditis elegans* (T. Johnson, personal communication). In some systems, stress-induced activation of the

ERK signalling pathway is also attenuated with ageing⁴³. Like the heat-shock response, ERK activation exerts a pro-survival signal during oxidative stress⁴⁴; thus, reduced ERK activity in aged cells may have a negative impact on survival. It is also worth noting that activation of ERK in response to mitogenic stimulation is also reduced as a function of ageing⁴⁵, again emphasizing the intimate link between proliferative and oxidative stress-response pathways. Basal DNA-binding activity of NF- κ B has been shown to increase with age, which again has been suggested to reflect increased oxidative stress in aged cells and tissues^{46,47}. However, as with heat-shock protein expression, evidence indicates that acute activation of this transcription factor by extracellular signals in T cells is diminished with ageing⁴⁸.

Genetics, oxidative phenomena and longevity

If oxidative stress and the ability to respond appropriately to it is important in ageing, then it follows that factors that increase resistance to stress should have anti-ageing benefits and lead to enhanced life span. In support of this claim, genetic links between stress responsiveness and longevity have been established in *C. elegans*, *Drosophila* and mice (Table 1). In *C. elegans*, such a relationship between life span and stress resistance was first demonstrated for *age-1* mutants which also displayed age-dependent elevations in CuZn-SOD and catalase activities^{49,50}. A variety of other life-extending mutations that are likewise correlated with enhanced stress tolerance have since been described (see Table 1 for references).

To the extent that their functions are known, many of the mutated genes encode proteins involved in regulating energy use, and therefore potentially the level of ROS. For example, the *age-1*, *daf-2* and *daf-16* genes in *C. elegans* are associated with an insulin-like signalling pathway that regulates dauer larval formation allowing the worm to survive periods of food scarcity. *age-1* and *daf-2* seem to suppress the activity of the downstream target *daf-16*, a forkhead transcription factor⁵¹. Hence, loss of function of either of these upstream regulators enhances *daf-16* function and leads to increased life span. Importantly, loss-of-function mutations in *daf-16* not only prevent longevity conferred by the *age-1* and *daf-2* mutations, but also abolish stress resistance⁵², thus strengthening the intimate link between longevity and stress responsiveness associated with this pathway. What remains to be determined is how the dauer mutations influence stress responsiveness and longevity. Identification of the transcriptional targets of *daf-16* should provide insight into this process. Another *C. elegans* gene whose mutation confers both increased longevity and enhanced stress responsiveness is *clk-1* (ref. 53). *clk-1* encodes a mitochondrial protein homologous to a yeast protein involved in the synthesis of coenzyme Q, an electron carrier required for respiration (Fig. 2). Although its biological function in the nematode is unclear, *clk-1* mutants are believed to enhance life span by reducing the rate of metabolism, which in turn might lead to slower accumulation of damage resulting from metabolic by-products such as ROS. Consistent with this hypothesis, overexpression of *clk-1* leads to a reduction in life span⁵⁴.

Gene mutations resulting in reduced life span in *C. elegans* have also been described. Although a life-shortening phenotype is generally considered to be less reliable than a life-lengthening trait for the assessment of ageing-associated genes, two such mutations, *mev-1* and *ctl-1*, are worth special mention as they seem to be linked to oxidative stress. *mev-1* encodes a subunit of the enzyme succinate dehydrogenase cytochrome *b*, a component of complex II of the mitochondrial electron transport chain⁵⁵. These animals have compromised mitochondrial function, reduced cytoplasmic SOD activity and show hypersensitivity to oxygen. *ctl-1* encodes a unique catalase localized in the cytosol⁵⁶. It is likely to contribute to the cell's antioxidant defences, but *ctl-1* mutants have not been examined directly for responsiveness to oxidative stress. Interestingly, loss of *ctl-1* function also eliminates longevity conferred by both the dauer pathway mutants and *clk-1*. Additional evidence for a link between

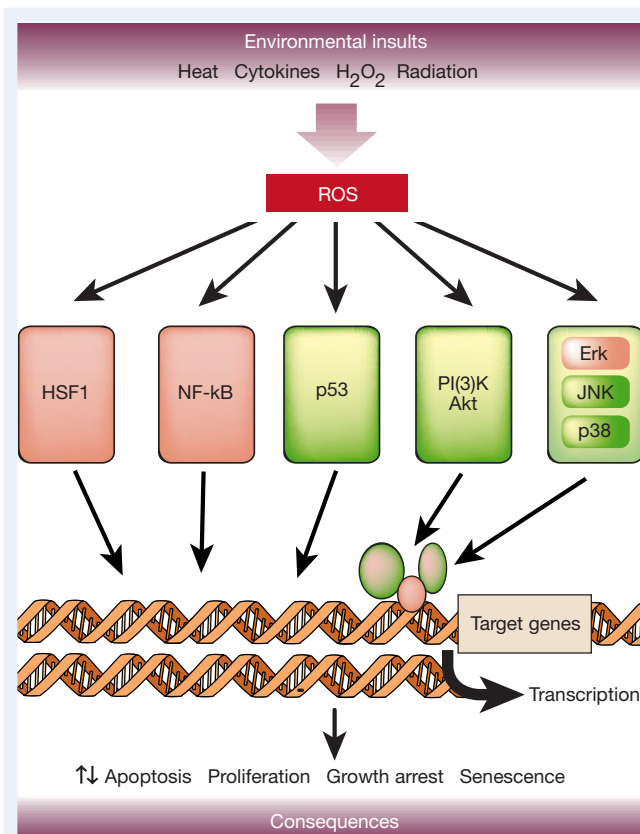


Figure 3 Major signalling pathways activated in response to oxidative stress. ROS can originate outside the cell, or may be generated intracellularly in response to external stimuli. Heat-shock transcription factor 1 (HSF1), NF- κ B and p53 are themselves transcription factors, while the PI(3)K/Akt and MAPK pathways regulate transcription factors through phosphorylation. The degree to which a given pathway is activated is highly dependent on the nature and duration of the stress, as well as the cell type. The consequences of the response vary widely, with the ultimate outcome being dependent on the balance between these stress-activated pathways. HSF1 is responsible for activation of the heat-shock response. Factors depicted in pink represent those pathways whose activities are altered with ageing.

longevity and oxidative stress resistance in the nematode has recently been obtained using a pharmacological approach to boost antioxidant defences⁵⁷. Treatment of wild-type *C. elegans* with synthetic SOD/catalase mimetics was shown to extend the mean life span by 44%. Moreover, the agent was also effective in restoring normal life span to *mev-1* mutants.

A similar link between longevity and stress resistance also exists in *Drosophila*. Various strains of flies selected for extended life span display increased resistance to oxidative stress that in some cases is also correlated with enhanced activity of antioxidant enzymes^{58,59}. In addition, at least one gene, *methuselah* (*mth*), has been identified whose mutation not only enhances longevity but also increases resistance to heat stress and paraquat (an intracellular ROS generator)⁶⁰. Further support for the relationship between stress tolerance and life span in *Drosophila* has been obtained using a transgenic approach to overexpress antioxidant genes. Overexpression of glutathione reductase, an enzyme involved in the generation of reduced glutathione, increases the resistance of flies exposed to hyperoxic conditions and extends life span under such conditions, although it does not affect life span of flies maintained in normal air⁶¹. Several studies have shown that elevated expression of CuZn-SOD alone or in combination with catalase likewise results in enhanced stress tolerance and increased life span⁶²⁻⁶⁴. However, other studies have reported little or no extension in longevity or enhanced stress

Box 1

Oxidants and diseases of ageing

Although maximum life span is the most relevant and defined endpoint with regard to ageing, it seems likely that in large multicellular organisms ageing need not proceed uniformly. This concept of non-uniform or focal ageing, or what has been termed segmental progeria by others⁸³, may be particularly important in a variety of age-related human diseases. Certainly the incidence of a host of diseases, including both cardiovascular and neurodegenerative disorders, increases exponentially with age. The basis for this steep rise in disease incidence is unexplained, but one possibility is that such diseases may share common mechanisms with ageing. In this regard, increasing evidence indicates that reactive oxygen species (ROS) may participate in the pathogenesis of these diseases. In support of this notion is the experimental evidence that the vessel wall of patients with atherosclerotic risk factors, but no overt disease, is characterized by a significant increase in vascular ROS production⁸⁶. In addition, many ophthalmological and neurodegenerative diseases seem to be mediated, at least in part, by oxidative stress.

One common feature to many of these conditions is the recruitment of inflammatory cells. These cells contribute to oxidative stress in large part because they contain the potent NADPH oxidase

system. Once activated, the NADPH oxidase complex, which is present in neutrophils, macrophages, microglia and vascular cells, produces large amounts of superoxide. In both the brain and the vessel wall the interaction of superoxide with endogenously generated nitric oxide can lead to the formation of peroxynitrite and other damaging radical species. Although the relative contribution of these various radical species is unknown, the link between inflammation and ROS seems to provide a useful framework for understanding disease progression.

Although laboratory data clearly suggest a role for oxidants in the pathogenesis of many age-related diseases, the clinical use of antioxidant therapy has been at best equivocal. Indeed, most of the large randomized trials of antioxidant vitamins such as vitamin A and vitamin E have shown, at best, marginal results and occasionally, potentially harmful outcomes (see Table below). Nonetheless, antioxidant therapy is in its infancy and issues of dosing and duration remain largely unexplored. It therefore is interesting to speculate that, in contrast to the relatively rare progerias that are characterized by accelerated global ageing, the clinical manifestations of these more common diseases, namely an atherosclerotic plaque or an Alzheimer tangle, may represent the fingerprint of accelerated but focal ageing.

Box 1 Table Oxidants, antioxidants and diseases of ageing

Disease system	Laboratory/animal studies	Clinical data
Cardiovascular	Pre-atherosclerotic blood vessels have increased levels of ROS ⁸⁶ Vitamin E protects against development of atherosclerosis ⁸⁷ Disruption of SOD leads to heart failure ^{88,89} and overexpression protects against injury ⁹⁰	PHS I: no overall benefit of beta-carotene on CVD? Benefit in high-risk subgroup ⁸⁹ CHAOS trial: vitamin E reduces rate of non-fatal myocardial infarct ¹⁰⁰ ATBC study: no overall benefit on CVD rate with vitamin E or beta-carotene? ¹ Increase in CVD deaths with beta-carotene ¹⁰
Ophthalmological	Offspring of pregnant mice depleted of glutathione develop cataracts ⁹¹ Retinal pigments produce ROS after light exposure ⁹² Retinal degeneration in primates with vitamin A or E deficiencies ⁹³	PHS I: non-significant reduction in cataracts and macular degeneration with vitamin E and multivitamins ¹⁰² NHS: carotenoids intake may decrease risk of cataracts ¹⁰³
Neurological	Mutations in SOD1 result in human ALS ⁹⁴ and transgenic animal models rescued by antioxidants ^{95,96} NMDA-receptor stimulation produces superoxide ⁹⁷ Defects in the function of complex 1 seen in Parkinson's disease ⁹⁸	Vitamin E not protective in early Parkinson's disease ¹⁰⁴ Vitamin E beneficial in Alzheimer's disease ¹⁰⁵ N-acetylcysteine does not effect survival in ALS ¹⁰⁶

The references cited above should be viewed as only representative examples derived from a much larger, relevant body of literature, which owing to space constraints cannot be fully presented. Acronyms and abbreviations: PHS I, Physicians' Health Study I; CHAOS, Cambridge Heart Antioxidant Study; ATBC, Alpha-Tocopherol, Beta-Carotene Cancer Prevention Study; NHS, Nurses' Health Study; CVD, cardiovascular disease; ALS, amyotrophic lateral sclerosis; NMDA, N-methyl-D-aspartate glutamate receptors.

resistance with overexpression of CuZn-SOD (refs 65, 66), and the basis for these discrepancies is presently unclear.

The effect of elevated SOD expression on stress resistance and life span has also been examined in mice. Although several reports have produced evidence to indicate that elevated CuZn-SOD levels afford protection against acute oxidative stress, a recent study showed that such acute protection did not correlate with increased longevity⁶⁷. However, another study found an unexpected correlation between life span and control of the oxidative stress response: targeted mutation of the *shc* gene locus, which caused ablation of a splice variant of relative molecular mass 66,000 (p66^{shc}), resulted in selective resistance to oxidative stress and extended life span⁶⁸. p66^{shc} is one of three splice variants derived from the *shc* locus. p52^{shc} and p46^{shc} are important adaptor proteins that are regulated by tyrosine phosphorylation and participate in growth-factor- and stress-induced ERK activation. p66^{shc} also undergoes tyrosine phosphorylation in response to extracellular signals, but its role in regulating ERK activation is less clear. However, the p66^{shc} variant is unique in that it also becomes phosphorylated on serine residues in response to certain extracellular signals and, in the case of oxidative stress, this leads to apoptosis. Serine phosphorylation of p66^{shc} is also stimulated by

insulin⁶⁹. This observation could provide another link to stress resistance and longevity, as an insulin-like pathway is clearly involved in regulating longevity and stress resistance in *C. elegans* (see above). Although better understanding of the mechanisms involved in p66^{shc}-mediated apoptosis and its importance to ageing await further investigation, these studies provide the first evidence for a genetic link between longevity and oxidative stress resistance in mammals.

Caloric restriction and oxidative stress

Limiting food intake, or caloric restriction, has been shown to extend life span in a wide range of species and in rodents it also slows the progression of a variety of age-associated diseases⁷⁰. Although several theories have been advanced over the years to explain the anti-ageing effects of caloric restriction, one favoured hypothesis proposes that it acts by decreasing oxidative stress⁷¹. In support of this hypothesis, the rate of oxidant generation of mitochondria from calorically restricted mice is significantly lower than from their ad libitum-fed counterparts, and caloric restriction reduces the age-associated accumulation of oxidatively damaged proteins, lipids and DNA⁷⁰. Caloric restriction also prevents many of the changes in gene expression and transcription-factor activity that normally occur with ageing,

including basal elevations in expression of heat-shock proteins^{40,41} and attenuation of stress-induced Hsp70 expression⁷². Finally, caloric restriction increases the ability of rodents to withstand a wide range of physiological stresses, improves thermotolerance and reduces heat-induced cellular damage in aged rats⁷³.

Therapeutic strategies against ageing

Understandably, there is tremendous public interest in the development of anti-ageing therapies. If one accepts the evidence that oxidative stress has a significant role in ageing processes, and that the ability to resist or prevent oxidative stress is a key determinant of longevity, then it is likely that strategies aimed either at reducing the oxidative burden or boosting host defence mechanisms involved in coping with the damage would have significant anti-ageing effects (see Box 1). Two key issues will then arise. First, can effective therapeutics be designed to combat oxidant damage? Second, would such approaches be effective in retarding ageing and/or preventing diseases associated with ageing? It is already known that caloric restriction fulfils both of these criteria in lower organisms, and experiments in non-human primates have produced encouraging results to suggest that reducing food intake might also delay ageing in humans⁷⁴. However, the possible application of caloric restriction to the human population would surely face insurmountable ethical and practical difficulties that would prevent its use. On the other hand, better understanding of the specific mechanisms responsible for the anti-ageing influences of caloric restriction could lead to the development of pharmacological agents (that is, caloric restriction mimetics) that would be effective in slowing ageing or delaying the onset of age-associated diseases.

Enhancement of antioxidant defences through dietary supplementation would seem to provide a more reasonable and practical approach to reduce the level of oxidative stress and there is a wealth of evidence to support the effectiveness of such a strategy *in vitro*. However, although some success has been achieved using antioxidant therapy in mammalian models of disorders induced by oxidative stress, so far these strategies have had little or no effect in enhancing longevity^{75,76}. Successful application of this approach will probably require much greater understanding of the pharmacological properties of many of the agents being used, including their rates of absorption, tissue distribution, metabolism and the microenvironment in which they must act. Indeed, the problem is clearly more complicated than simply adding a pharmacological agent, as most free-radical scavengers act in oxidation-reduction reactions that are reversible, and some, such as ascorbate, can act both as antioxidants and pro-oxidants, depending on the conditions⁷⁷. In addition, given the role of ROS as mediators of normal signalling processes, determination of the optimal dosage of supplements may require fine tuning to avoid overshooting the desired effects to the point of perturbing the delicate redox balance required for the maintenance of normal cell functions (Fig. 1). However, the development of new synthetic compounds that act as mimetics of SOD and catalase has offered an alternative approach with some promise^{78,79}. Use of such compounds in mouse models has been effective in attenuating oxidative stress-associated disease processes, and leads to extension of longevity in *C. elegans*⁵⁷.

Finally, consideration should be given to strategies to boost the host defence mechanisms that are known to be activated in response to oxidative stress. So far, the best mechanism for boosting such responses seems to be stress itself. That is, a sublethal or conditioning stress can lead to enhanced survival and reduced tissue damage following a subsequent, more severe stress. This concept, termed hormesis^{39,70,71}, is in many ways the physiological equivalent of the philosophical notion that 'what won't kill you, will make you strong'. Although the basis for this stress tolerance phenomenon is poorly understood, elevations in heat-shock proteins may be one important factor. Interestingly, such conditioning stress can also lead to life extension in both *C. elegans*^{80,81} and *Drosophila*⁸². In humans as well as

lower mammals, exercise is believed to have many anti-ageing benefits. Because even in conditioned athletes exercise results in acute stress, it is possible that some of its benefits are derived through such a stress-tolerance mechanism.

Conclusions

In summary, although ageing is likely to be a multifactorial process, there is now significant evidence implicating the generation of ROS and the corresponding response to oxidative stress as key factors in determining longevity. Much of the early evidence was correlative, such as the relationship between metabolic rate and life span, and the observed increase in oxidative damage as a function of age. However, the more recent identification of longevity-influencing genes, first in lower organisms and subsequently in mammals, significantly strengthens the mechanistic connection between oxidants, stress and ageing. Many questions remain including how these various longevity-associated genes actually function to influence stress resistance and ageing, and how applicable they are to human ageing. If oxidative stress does indeed contribute to ageing in the manner described, do ROS act purely as random, destructive agents or — as we have intimated in this review — as regulators of discrete pathways linked to stress responsiveness and ageing? Similarly, it remains unclear whether it is the absolute level of oxidative stress encountered or the subsequent appropriate or inappropriate response to oxidative stress, or perhaps some combination of both, that ultimately determines life expectancy. Finally, and very importantly, in what fashion does oxidative stress intersect, if at all, with other proposed determinants of ageing such as telomere length and telomerase activity, or gene products such as the Werner's helicase that are linked to accelerated ageing syndromes? Nearly a century after the rate-of-living hypothesis was first proposed, the answers to these and related questions remain largely unknown. However, given the rapid rate at which our knowledge in this area has increased over the past few years, it is likely that answers to many of these questions will be forthcoming within this first decade of the new century. These answers will undoubtedly help determine whether ROS are merely peripheral targets that correlate with longevity, or instead, are finally established as a central regulator of human ageing. □

1. Harman, D. Aging: a theory based on free radical and radiation chemistry. *J. Gerontol.* **2**, 298–300 (1957).
2. McCord, J. M. & Fridovich, I. Superoxide dismutase. An enzymatic function for erythrocuperin (hemocuperin). *J. Biol. Chem.* **244**, 6049–6055 (1969).
3. Ku, H. H., Brunk, U. T. & Sohal, R. S. Relationship between mitochondrial superoxide and hydrogen peroxide production and longevity of mammalian species. *Free Radical Biol. Med.* **15**, 621–627 (1993).
4. Finkel, T. Oxygen radicals and signaling. *Curr. Opin. Cell Biol.* **10**, 248–253 (1998).
5. Nishikawa, T. *et al.* Normalizing mitochondrial superoxide production blocks three pathways of hyperglycaemic damage. *Nature* **404**, 787–790 (2000).
6. Nemoto, S., Takeda, K., Yu, Z. X., Ferrans, V. J. & Finkel, T. A role for mitochondrial oxidants as regulators of cellular metabolism. *Mol. Cell. Biol.* **20**, 7311–7318 (2000).
7. Suh, Y. A. *et al.* Cell transformation by the superoxide-generating oxidase Mox1. *Nature* **401**, 79–82 (1999).
8. Geiszt, M., Kopp, J. B., Várnai, P., & Leto, T. L. Identification of Renox, an NAD(P)H oxidase in kidney. *Proc. Natl Acad. Sci. USA* **97**, 8010–8014 (2000).
9. Turrens, J. F. Superoxide production by the mitochondrial respiratory chain. *Biosci. Rep.* **17**, 3–8 (1997).
10. Golubev, A. G. The other side of metabolism: a review. *Biochemistry* **61**, 1443–1460 (1996).
11. Boveris, A. & Chance, B. The mitochondrial generation of hydrogen peroxide. General properties and effect of hyperbaric oxygen. *Biochem. J.* **134**, 707–716 (1973).
12. Skulachev, V. P. Role of uncoupled and non-coupled oxidations in maintenance of safely low levels of oxygen and its one-electron reductants. *Q. Rev. Biophys.* **29**, 169–202 (1996).
13. Vidal-Puig, A. J. *et al.* Energy metabolism in uncoupling protein 3 gene knockout mice. *J. Biol. Chem.* **275**, 16258–16266 (2000).
14. Chae, H. Z., Kang, S. W. & Rhee, S. G. Isoforms of mammalian peroxiredoxin that reduce peroxides in presence of thioredoxin. *Methods Enzymol.* **300**, 219–226 (1999).
15. Stadtman, E. R. Protein oxidation and aging. *Science* **257**, 1220–1224 (1992).
16. Yan, L. J., Levine, R. L. & Sohal, R. S. Oxidative damage during aging targets mitochondrial aconitase. *Proc. Natl Acad. Sci. USA* **94**, 11168–11172 (1997).
17. Goto, S. *et al.* Carbonylated proteins in aging and exercise: immunoblot approaches. *Mech. Age Dev.* **107**, 245–253 (1999).
18. Dukan, S. *et al.* Protein oxidation in response to increased transcriptional or translational errors. *Proc. Natl Acad. Sci. USA* **97**, 5746–5749 (2000).
19. Beckman, K. B. & Ames, B. N. The free radical theory of aging matures. *Physiol. Rev.* **78**, 547–581 (1998).

20. Esposito, L. A., Melov, S., Panov, A., Cottrell, B. A. & Wallace, D. C. Mitochondrial disease in mouse results in increased oxidative stress. *Proc. Natl Acad. Sci. USA* **96**, 4820–4825 (1999).
21. Melov, S. *et al.* Mitochondrial disease in superoxide dismutase 2 mutant mice. *Proc. Natl Acad. Sci. USA* **96**, 846–851 (1999).
22. Packer, L. & Fuehr, K. Low oxygen concentration extends the lifespan of cultured human diploid cells. *Nature* **267**, 423–425 (1977).
23. von Zglinicki, T., Saretzki, G., Docke, W. & Lotze, C. Mild hyperoxia shortens telomeres and inhibits proliferation of fibroblasts: a model for senescence? *Exp. Cell Res.* **220**, 186–193 (1995).
24. Chen, Q. & Ames, B. N. Senescence-like growth arrest induced by hydrogen peroxide in human diploid fibroblast F65 cells. *Proc. Natl Acad. Sci. USA* **91**, 4130–4134 (1994).
25. Serrano, M., Lin, A. W., McCurrach, M. E., Beach, D. & Lowe, S. W. Oncogenic ras provokes premature cell senescence associated with accumulation of p53 and p16INK4a. *Cell* **88**, 593–602 (1997).
26. Lee, A. C. *et al.* Ras proteins induce senescence by altering the intracellular levels of reactive oxygen species. *J. Biol. Chem.* **274**, 7936–7940 (1999).
27. Johnson, T. M., Yu, Z. X., Ferrans, V. J., Lowenstein, R. A. & Finkel, T. Reactive oxygen species are downstream mediators of p53-dependent apoptosis. *Proc. Natl Acad. Sci. USA* **93**, 11848–11852 (1996).
28. Polyak, K., Xia, Y., Zweier, J. L., Kinzler, K. W. & Vogelstein, B. A model for p53-induced apoptosis. *Nature* **389**, 300–305 (1997).
29. Sachsenmaier, C. *et al.* Involvement of growth factor receptors in the mammalian UVC response. *Cell* **78**, 963–972 (1994).
30. Wang, X., McCullough, K. D., Franke, T. F. & Holbrook, N. J. Epidermal growth factor receptor-dependent Akt activation by oxidative stress enhances cell survival. *J. Biol. Chem.* **275**, 14624–14631 (2000).
31. Knebel, A., Rahmsdorf, H. J., Ullrich, A. & Herrlich, P. Dephosphorylation of receptor tyrosine kinases as target of regulation by radiation, oxidants or alkylating agents. *EMBO J.* **15**, 5314–5325 (1996).
32. Lee, S. R., Kwon, K. S., Kim, S. R. & Rhee, S. G. Reversible inactivation of protein-tyrosine phosphatase 1B in A431 cells stimulated with epidermal growth factor. *J. Biol. Chem.* **273**, 15366–15372 (1998).
33. Burdon, R. H. Superoxide and hydrogen peroxide in relation to mammalian cell proliferation. *Free Radical Biol. Med.* **18**, 775–794 (1995).
34. Saitoh, M. *et al.* Mammalian thioredoxin is a direct inhibitor of apoptosis signal-regulating kinase (ASK) 1. *EMBO J.* **17**, 2596–2606 (1998).
35. Adler, V. *et al.* Regulation of JNK signaling by GSTp. *EMBO J.* **18**, 1321–1334 (1999).
36. Allen, R. G. & Tresini, M. Oxidative stress and gene regulation. *Free Radical Biol. Med.* **28**, 463–499 (2000).
37. Godon, C. *et al.* The H₂O₂ stimulant in *Saccharomyces cerevisiae*. *J. Biol. Chem.* **273**, 22480–22489 (1998).
38. Richardson, A. A. & Holbrook, N. J. In *Cellular Aging and Cell Death* (eds Holbrook, N. J., Martin, G. R. & Lockshin, R. A.) 67–89 (Wiley, New York, 1996).
39. Morimoto, R. I. & Santoro, M. G. Stress-inducible responses and heat shock proteins: new pharmacologic targets for cytoprotection. *Nature Biotechnol.* **16**, 833–838 (1998).
40. Lee, C. K., Klopp, R. G., Weindrich, R. & Prolla, T. A. Gene expression profile of aging and its retardation by caloric restriction. *Science* **285**, 1390–1393 (1999).
41. Lee, C. K., Weindrich, R. & Prolla, T. A. Gene-expression profile of the ageing brain in mice. *Nature Genet.* **25**, 294–297 (2000).
42. Wheeler, J. C., Bieschke, E. T. & Tower, J. Muscle-specific expression of *Drosophila* hsp70 in response to aging and oxidative stress. *Proc. Natl Acad. Sci. USA* **92**, 10408–10412 (1995).
43. Guyton, K. Z. *et al.* Age-related changes in activation of mitogen-activated protein kinase cascades by oxidative stress. *J. Invest. Dermatol. Symp. Proc.* **3**, 23–27 (1998).
44. Guyton, K. Z. *et al.* Activation of mitogen-activated protein kinase by hydrogen peroxide: role in cell survival following oxidant injury. *J. Biol. Chem.* **271**, 4138–4142 (1996).
45. Liu, *et al.* Age-related decline in MAP kinase activity in EGF-stimulated rat hepatocytes. *J. Biol. Chem.* **271**, 3604–3607 (1996).
46. Poynter, M. E. & Daynes, R. A. Peroxisome proliferator-activated receptor alpha activation modulates cellular redox status, represses nuclear factor-kappaB signaling, and reduces inflammatory cytokine production in aging. *J. Biol. Chem.* **273**, 32833–32841 (1998).
47. Supakar, P. C., Jung, M. H., Song, C. S., Chatterjee, B. & Roy, A. K. Nuclear factor kappa B functions as a negative regulator for the rat androgen receptor gene and NF-kappa B activity increases during the age-dependent desensitization of the liver. *J. Biol. Chem.* **270**, 837–842 (1995).
48. Ponnappan, U. Regulation of transcription factor NF-kappa B in immune senescence. *Front. Biosci.* **1**, D152–D168 (1998).
49. Larsen, P. L. Aging and resistance to oxidative damage in *Caenorhabditis elegans*. *Proc. Natl Acad. Sci. USA* **90**, 8905–8909 (1993).
50. Vanfleteren, J. R. Oxidative stress and ageing in *Caenorhabditis elegans*. *Biochem. J.* **292**, 605–608 (1993).
51. Ogg, S. *et al.* The Fork head transcription factor DAF-16 transduces insulin-like metabolic and longevity signals in *C. elegans*. *Nature* **389**, 994–999 (1997).
52. Murakami, S. & Johnson, T. E. A genetic pathway conferring life extension and resistance to UV stress in *Caenorhabditis elegans*. *Genetics* **143**, 1207–1218 (1996).
53. Wong, A., Boutis, P. & Hekimi, S. Mutations in the clk-1 gene of *Caenorhabditis elegans* affect developmental and behavioral timing. *Genetics* **139**, 1247–1259 (1995).
54. Branicky, R., Benard, C. & Hekimi, S. clk-1, mitochondria, and physiological rates. *Bioessays* **22**, 48–56 (2000).
55. Ishii, N. *et al.* A mutation in succinate dehydrogenase cytochrome b causes oxidative stress and ageing in nematodes. *Nature* **394**, 694–697 (1998).
56. Taub, J. *et al.* A cytosolic catalase is needed to extend adult lifespan in *C. elegans* *daf-1* and *clk-1* mutants. *Nature* **399**, 162–166 (1999).
57. Melov, S. *et al.* Extension of lifespan with superoxide dismutase/catalase mimetics. *Science* **289**, 1567–1569 (2000).
58. Dudas S. *et al.* A coordinate upregulation of antioxidant gene activities is associated with the delayed onset of senescence in a long-lived strain of *Drosophila*. *J. Gerontol. A* **50**, B117–B127 (1995).
59. Harshman, L. G. & Haberer, B. A. Oxidative stress resistance: a robust correlated response to selection in extended longevity lines of *Drosophila melanogaster*. *J. Gerontol. A* **55**, B415–B417 (2000).
60. Lin, Y. J., Seroude, L. & Benzer, S. Extended life-span and stress resistance in the *Drosophila* mutant methuselah. *Science* **282**, 943–946 (1998).
61. Mockett, R. J., Sohal, R. S. & Orr, W. C. Overexpression of glutathione reductase extends survival in transgenic *Drosophila melanogaster* under hyperoxia but not normoxia. *FASEB J.* **13**, 1733–1742 (1999).
62. Orr, W. C. & Sohal, R. S. Extension of life-span by overexpression of superoxide dismutase and catalase in *Drosophila melanogaster*. *Science* **263**, 1128–1130 (1994).
63. Sun, J. & Tower, J. FLP recombinase-mediated induction of Cu/Zn-superoxide dismutase transgene expression can extend the life span of adult *Drosophila melanogaster* flies. *Mol. Cell Biol.* **19**, 216–228 (1999).
64. Parkes, T. L. *et al.* Extension of *Drosophila* lifespan by overexpression of human SOD1 in motoneurons. *Nature Genet.* **19**, 171–174 (1998).
65. Seto, N. O., Hayashi, S. & Tener, G. M. Overexpression of Cu-Zn superoxide dismutase in *Drosophila* does not affect life-span. *Proc. Natl Acad. Sci. USA* **87**, 4270–4274 (1990).
66. Orr, W. C. & Sohal, R. S. Effects of Cu-Zn superoxide dismutase overexpression on life span and resistance to oxidative stress in transgenic *Drosophila melanogaster*. *Arch. Biochem. Biophys.* **301**, 34–40 (1993).
67. Huang, T. *et al.* Ubiquitous overexpression of CuZn superoxide dismutase does not extend life span in mice. *J. Gerontol. Biol. Sci. A* **55**, B5–B9 (2000).
68. Migliaccio, E. *et al.* The p66^{shc} adaptor protein controls oxidative stress response and life span in mammals. *Nature* **402**, 309–313 (1999).
69. Kao, A. W. *et al.* Insulin stimulates the phosphorylation of the 66- and 52-kilodalton Shc isoforms by distinct pathways. *Endocrinology* **138**, 2474–2480 (1997).
70. Masoro, E. J. Caloric restriction and aging: an update. *Exp. Gerontol.* **35**, 299–305 (2000).
71. Sohal, R. S. & Weindrich, R. Oxidative stress, caloric restriction, and aging. *Science* **273**, 59–63 (1996).
72. Heydari, A. R., Wu, B., Takahashi, R., Strong, R. & Richardson, A. Expression of heat shock protein 70 is altered by age and diet at the level of transcription. *Mol. Cell Biol.* **13**, 2909–2918 (1993).
73. Hall, D. M. *et al.* Caloric restriction improves thermotolerance and reduces hyperthermia-induced cellular damage in old rats. *FASEB J.* **14**, 78–86 (2000).
74. Roth, G. S., Ingram, D. K. & Lane, M. A. Calorie restriction in primates: will it work and how will we know? *J. Am. Geriatr. Soc.* **47**, 896–903 (1999).
75. McCall, M. R. & Frei, B. Can antioxidant vitamins maternally reduce oxidative damage in humans? *Free Radical Biol. Med.* **26**, 1034–1053 (1999).
76. Yu, B. P. Approaches to anti-aging intervention: the promises and the uncertainties. *Mech. Ageing Dev.* **111**, 73–87 (1999).
77. Podmore, I. D. *et al.* Vitamin C exhibits pro-oxidant properties. *Nature* **392**, 559–559 (1998).
78. Melov, S. *et al.* A novel neurological phenotype in mice lacking mitochondrial manganese superoxide dismutase. *Nature Genet.* **18**, 159–163 (1998).
79. Rong, Y. *et al.* EUK-134, a synthetic superoxide dismutase and catalase mimetic, prevents oxidative stress and attenuates kainate-induced neuropathology. *Proc. Natl Acad. Sci. USA* **96**, 9897–9902 (1999).
80. Lithgow, G. J. *et al.* Thermotolerance and extended life-span conferred by single-gene mutations and induced by thermal stress. *Proc. Natl Acad. Sci. USA* **92**, 7540–7544 (1995).
81. Johnson, T. E. & Hartman, P. S. Radiation effects on life span in *Caenorhabditis elegans*. *J. Gerontol. Biol. Sci.* **43**, B137–B141 (1988).
82. Tatar, M., Khazaei, A. A. & Curtsinger, J. W. Chaperoning extended life. *Nature* **390**, 30 (1997).
83. Martin, G. M., Austad, S. N. & Johnson, T. E. Genetic analysis of ageing: role of oxidative damage and environmental stresses. *Nature Genet.* **13**, 25–34 (1996).
84. Cypser, J. R. & Johnson, T. E. The *spe-10* mutant has longer life and increased stress resistance. *Neurobiol. Aging* **20**, 503–512 (1999).
85. Murakami, S. & Johnson, T. E. Life extension and stress resistance in *Caenorhabditis elegans* modulated by the *tkr-1* gene. *Curr. Biol.* **8**, 1091–1094 (1998).
86. Abe, J. & Berk, B. C. Reactive oxygen species as mediators of signal transduction in cardiovascular disease. *Trends Cardiovasc. Med.* **8**, 59–64 (1998).
87. Wojcicki, J. *et al.* Effect of selenium and vitamin E on the development of experimental atherosclerosis in rabbits. *Atherosclerosis* **87**, 9–16 (1991).
88. Lebovitz, R. M. *et al.* Neurodegeneration, myocardial injury, and perinatal death in mitochondrial superoxide dismutase-deficient mice. *Proc. Natl Acad. Sci. USA* **93**, 9782–9787 (1996).
89. Li, Y. B. *et al.* Dilated cardiomyopathy and neonatal lethality in mutant mice lacking manganese superoxide-dismutase. *Nature Genet.* **11**, 376–381 (1995).
90. Wang, P. H. *et al.* Overexpression of human copper, zinc-superoxide dismutase (SOD1) prevents postischemic injury. *Proc. Natl Acad. Sci. USA* **95**, 4556–4560 (1998).
91. Martenson, J., Steinherz, R., Jain, A. & Meister, A. Glutathione ester prevents buthionine sulfoximine-induced cataracts and lens epithelial cell damage. *Proc. Natl Acad. Sci. USA* **86**, 8727–8731 (1989).
92. Rozanowska, M. *et al.* Blue light-induced reactivity of retinal age pigment. In vitro generation of oxygen-reactive species. *J. Biol. Chem.* **270**, 18825–18830 (1995).
93. Hayes, K. C. Retinal degeneration in monkeys induced by deficiencies of vitamin E or A. *Invest. Ophthalmol.* **13**, 499–510 (1974).
94. Rosen, D. R. *et al.* Mutations in Cu/Zn superoxide dismutase gene are associated with familial amyotrophic lateral sclerosis. *Nature* **362**, 59–62 (1993).
95. Andreassen, O. A., Dedeglu, A., Klivenyi, P., Beal, M. F. & Bush, A. I. N-acetyl-L-cysteine improves survival and preserves motor performance in an animal model of familial amyotrophic lateral sclerosis. *Neuroreport* **11**, 2491–2493 (2000).
96. Nagano, S., Ogawa, Y., Yanagihara, T., & Sakoda, S. Benefit of a combined treatment with trientine and ascorbate in familial amyotrophic lateral sclerosis model mice. *Neurosci. Lett.* **265**, 159–162 (1999).
97. Lafon-Cazal, M., Pietri, S., Culcasi, M. & Bockaert, J. NMDA-dependent superoxide production and neurotoxicity. *Nature* **364**, 535–537 (1993).
98. Parker, W. D. Jr, Boyson, S. J. & Parks, J. K. Abnormalities of the electron transport chain in idiopathic Parkinson's disease. *Ann. Neurol.* **26**, 719–723 (1989).
99. Hennekens, C. H. *et al.* Lack of effect of long-term supplementation with beta carotene on the incidence of malignant neoplasms and cardiovascular disease. *N. Engl. J. Med.* **334**, 1145–1149 (1996).

100. Stephens, N. G. *et al.* Randomised controlled trial of vitamin E in patients with coronary disease. Cambridge Heart Antioxidant Study (CHAOS). *Lancet* **347**, 781–786 (1996).
101. The Alpha-Tocopherol Beta Carotene Study Group. The effect of vitamin E and beta carotene on the incidence of lung cancer and other cancers in male smokers. *N. Engl. J. Med.* **330**, 1029–1035 (1994).
102. Seddon, J. M. *et al.* The use of vitamin supplements and the risk of cataract among US male physicians. *Am. J. Public Health* **84**, 788–792 (1994).
103. Hankinson, S. E. *et al.* Nutrient intake and cataract extraction in women: a prospective study. *Br. Med. J* **305**, 335–339 (1992).
104. The Parkinson Study Group. Effects of tocopherol and deprenyl on the progression of disability in early Parkinson's disease. *N. Engl. J. Med.* **328**, 176–183 (1993).
105. Sano, M. *et al.* A controlled trial of selegiline, alpha-tocopherol, or both as treatment for Alzheimer's disease. The Alzheimer's Disease Cooperative Study. *N. Engl. J. Med.* **336**, 1216–1222 (1997).
106. Louwse, E. S., Weverling, G. J., Bossuyt, P. M., Meyjes, F. E. & de Jong, J. M. Randomized, double-blind, controlled trial of acetylcysteine in amyotrophic lateral sclerosis. *Arch. Neurol.* **52**, 559–564 (1995).

Acknowledgements

We thank I. Rovira and R. Wange for their help in the design of illustrations, and T. Johnson, D. Longo, B. Howard, R. Levine, M. Gorospe and N. Epstein for thoughtful comments and discussions.

The age of cancer

Ronald A. DePinho

Departments of Adult Oncology, Medicine and Genetics, Dana-Farber Cancer Institute, Harvard Medical School, Boston, Massachusetts 02115, USA (e-mail: ron_depinho@dfci.harvard.edu)

A striking link exists between advanced age and increased incidence of cancer. Here I review how several of the age-related molecular and physiological changes might act in concert to promote cancer, and in particular epithelial carcinogenesis. Experimental data indicate that the aged, cancer-prone phenotype might represent the combined pathogenetic effects of mutation load, epigenetic regulation, telomere dysfunction and altered stromal milieu. Further verification of the role of these effects should in turn lead to the design of effective therapeutics for the treatment and prevention of cancer in the aged.

Advancing age is the most potent of all carcinogens. In humans, the incidence of cancer rises exponentially in the final decades of life, culminating in a lifetime risk of 1 in 2 for men and 1 in 3 for women¹. This dramatic age-dependent escalation in cancer risk is fuelled largely by a marked increase in epithelial carcinomas from ages 40 to 80 years, as opposed to cancers of mesenchymal or haematopoietic origin (Fig. 1). What underlies this intimate link between cancer and advanced age? At first glance, one may presume that the sequential accumulation of somatic mutations over a lifetime finally drives a subset of cells in the organism over a critical threshold, leading to emergence of cancers later in life. If so, then why do epithelial carcinomas predominate? Even more perplexing is the fact that most laboratory mouse strains approaching a biologically equivalent age succumb instead to mesenchymal and haematopoietic malignancies. Here I consider the thesis that a number of molecular and physiological processes underlie the rate and patterns of cancer in the aged. Specifically, I propose that the cancer-prone phenotype of older humans might reflect the combined effects of cumulative mutational load, increased epigenetic gene silencing, telomere dysfunction and altered stromal milieu.

Genome maintenance in ageing mice and humans

Mutation rates and cancer

Cancer is clearly a disease of the genes². Tumour progression is driven by clonal selection and evolution of tumour cell populations³, a model well substantiated by serial mutational analyses of staged human colorectal cancers⁴ as well as other cancer types^{5–8}. An increase in somatic mutations has been documented in aged cells and tissues of both humans and mice, and presumably relates to cumulative lifetime exposure to endogenous and exogenous DNA damaging agents^{9,10}, as well as to an accumulation of mutations resulting from proof-reading and mismatch errors during DNA replication¹¹. Less clear is whether this process is exacerbated by a progressive decline in DNA monitoring and repair capability. The fundamental question remains whether the rates of spontaneous mutation *in vivo* are of sufficient magnitude to initiate and drive the transformation process and ultimately generate the wholesale genomic alterations encountered in adult human cancers. As a corollary, one wonders whether mutational rates differ sufficiently among various tissues to provide a basis for the observed tumour spectrum of later life.

It has been estimated that, in cultured human cells, the

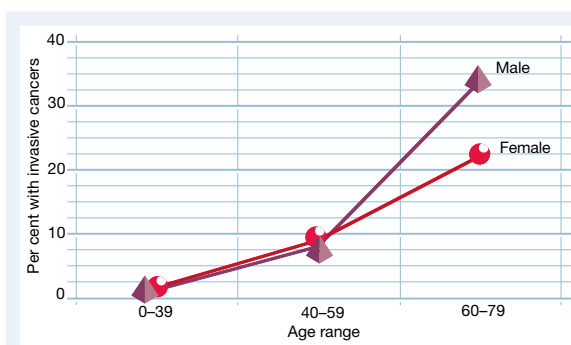


Figure 1 Cancer incidence as a function of age. Incidence of invasive cancer plotted against age ranges reveals exponential increase from age 40–80 years (ref. 1). Purple, male; red, female. Note that beyond age 80, incidence of cancers plateau⁹⁴.

spontaneous mutation rate is approximately 2×10^{-7} per gene per cell division¹². If this rate remains fixed over a lifetime and represents an accurate measure of *in vivo* mutational events across the intact organism, then each cell would accumulate only a few mutations over a lifetime¹³. This rate would be well below the predicted value of between four and ten rate-limiting, stochastic events for tumour initiation and progression, which was extrapolated from cancer trends in ageing populations and from transgenic models of tumorigenesis^{14–16}. The observation that rates of spontaneous mutations seem insufficient to account for the extensive tumour-associated genomic changes led to the concept, and subsequent proof, of the ‘mutator phenotype’^{17,18}. In this scenario, the age-dependent increase in cancer stems from the eventual mutation of genes governing genome stability, mutational inactivation of which would lead to an accelerated pace of mutations overall¹⁷. The spontaneous loss of function, or perhaps haploinsufficiency, in these genome maintenance pathways might also provide a basis for the well-recognized inherent instability of tumour cell genomes relative to their normal counterparts^{19,20}. Perhaps the disconnect between mutation rates and cancer genome profiles can be explained in part by the clonal expansion of cancerous stem cells harbouring initiating lesions. This basic idea is a reasonable one, although there is evidence that robust checkpoint mechanisms could prevent further expansion of nascent cancer cells harbouring oncogenic lesions, for example, Ras-induced senescence or Myc-induced apoptosis²¹. Thus, although mutator and clonal expansion mechanisms are likely to contribute to the

increase in cancer as a function of age, these mechanisms have not yet been documented in the context of ageing tissues and fail to provide a unifying principle that accounts for the tissue distribution and cytogenetic profiles of most adult cancers.

An alternative view proposes that mutation rates calculated on the basis of biological age rather than number of cell divisions could more closely approximate the requisite number of mutations presumed necessary for initiation of tumorigenesis²². Such rates are in accord with those empirically determined in mice harbouring an integrated mutational reporter^{23,24}. These model systems, reviewed by Vijg and Dolle²⁵, have revealed tissue-specific differences in mutational rates (for example, intestine and liver > heart > brain) as well as radically different mutational spectra in these aged tissues, the latter pointing to organ-specific differences in genome maintenance mechanisms. However, these studies indicate that age-dependent rates of mutation do not correlate strictly with rates of cell proliferation nor do they readily explain the observed tumour spectra of aged individuals.

Age-dependent deterioration in genome maintenance

Hereditary cancer and progeric syndromes (the latter characterized by physical symptoms suggestive of premature senility) have revealed key cancer suppression pathways, leading to the discovery of critical genome stability genes participating in processes of DNA repair, DNA replication, cell-cycle checkpoints and chromosome structural maintenance/segregation (see review in this issue by Martin and Oshima, pages 263–266). Many of these genes encode DNA polymerases and helicases as well as core components of the machinery for DNA mismatch repair, base excision repair, transcription coupled repair and non-homologous end-joining. The increased incidence of cancer associated with germ-line inactivation of the Werner helicase (WRN) and related helicases is particularly noteworthy as it underscores the inter-relationship of ageing, genome maintenance and cancer.

Curtis and Crowley²⁶ were the first to provide experimental evidence for an increasing level of genomic abnormalities as a function of age, reporting an increased number of abnormal metaphases in older, hepatectomized mice compared with younger animals (75% versus 10% aberrant metaphases, respectively). Similarly, aneuploidy, translocations and end-to-end fusions (dicentrics) have been shown to be higher in peripheral blood lymphocytes and fibroblasts of elderly humans compared with younger individuals^{27–29}. An increase in chromosomal anomalies has also been observed in ageing mice, but the number of chromosomal dicentrics remained low throughout life³⁰, a finding that probably relates to species differences in telomere dynamics (see below). From these studies, it is clear that the process of normal ageing *per se* is associated with an overall deterioration in genome integrity, which might reflect an age-related decline in repair capabilities, or merely imperfect repair in the setting of age-dependent accumulation of mutations.

Several groups have provided some experimental data in support of the existence of declining repair capacity. Specifically, an age-related decline in the ability to process new ultraviolet (UV) light-induced DNA damage (measured by the ability to repair a UV-treated reporter plasmid) has been demonstrated in cultured primary skin fibroblasts and lymphoblastoid cell lines from normal donors in their first to tenth decade of life³¹. This age-associated decrease in the repair of UV-induced DNA damage coincides with decreased levels of proteins that participate in the repair process, such as ERCC3 (for excision repair cross-complementing 3), PCNA (for proliferating cell nuclear antigen), RPA (for replication protein A) and p53 (ref. 32), and with increased incidence of skin cancer in the aged. However, others have found no significant age-related differences in repair-replication response of human epidermal keratinocytes to UV-induced DNA damage³³. The lack of an assay for the quantitative measurement of *in vivo* repair in humans has made this a difficult question to answer definitively.

Regardless, age-dependent changes in DNA repair proficiency have not been established unequivocally as a mechanism driving

carcinogenesis in the aged. However, germ-line defects in mismatch repair can drive late-onset epithelial carcinogenesis in a subset of the general population. This has become apparent by the recognition that germ-line MutS homologue 6 (MSH6) mutations (a less critical component of the mismatch repair complex) are associated with colon cancers, although their late onset had led to the erroneous assignment of these hereditary cancers as 'sporadic'³⁴. Although it is possible that genomic screens of sporadic adult cancers will uncover additional examples of 'low-penetrant' DNA repair alleles, most colon cancers as well as other carcinoma types do not exhibit the signature 'MIN' genomic profiles (that is, microsatellite instability resulting from mismatch repair deficiency) associated with DNA repair defects¹⁹, implying that other mechanisms exist to drive epithelial carcinogenesis in the aged.

Epigenetic mechanisms and carcinogenesis

The lack of complete concordance between mutational rates and cancer profiles in the aged indicates that gene function and/or regulation might be affected by means of a non-structural mechanism. One class, termed epigenetic mechanisms, have been shown to operate primarily on the levels of DNA methylation and chromatin structure. Growing evidence exists that these epigenetic mechanisms modulate many different genes in an age-dependent and tissue-specific manner. *De novo* hypermethylation of CpG-rich islands (regions rich in cytosine–guanine doublets) nested in gene promoters has been shown to be a common means of silencing tumour-suppressor genes in cancers^{35,36}. The observation that age-progressive CpG-island methylation takes place in a subset of cells residing in normal tissues and seems to operate in a gene-specific manner raises the possibility of targeting cancer-relevant pathways³⁷. Issa and colleagues have proposed that this rising tide of promoter methylation in such genes might presage the intensification of methylation of these target sequences coincident with progression towards full malignant transformation³⁸. That some of these methylated genes (for example, oestrogen receptor) have plausible links to growth control and oncogenesis has fuelled speculation that age-progressive methylation contributes to the silencing of cancer-relevant genes and hence to increased cancer incidence in the aged.

This hypothesis is supported by the genetic observation that the net growth rate and multiplicity of intestinal adenomas in the Min mouse (germ-line *APC* mutation) are reduced with loss of one copy of the prototypic DNA methyltransferase, Dnmt1 (refs 39, 40). On the other hand, the somatic disruption of DNMT1 in cultured human colorectal carcinoma cells, although associated with an overall decrease in genomic methylation, did not prevent methylation and silencing of many loci including the p16INK4a tumour-suppressor gene, thus pointing to the existence of other DNA methylating activities and their regional specificity of action⁴¹. However, the impact of DNMT1 disruption, or more generally of reduced genomic methylation, in pre-neoplastic cells might be very different from that in transformed cells. A detailed view of the molecular machinery responsible for these gene-specific methylation patterns and how the activity of this machinery is regulated in ageing tissues remain critical challenges for the future.

Telomere dysfunction and epithelial carcinogenesis

Age-dependent genetic and epigenetic events are likely contributors to the increased incidence of cancer in later life. Less evident is how such processes spur the preferential development of epithelial cancers in older humans, while sarcomas and lymphomas predominate in the paediatric population (and in mice at any age). Known tissue- and species-specific differences in the rates and types of somatic mutation seem insufficient to account for these observed cancer patterns. On a similar note, it is not clear whether differences in the relative prominence of key tumour-suppressor and repair pathways in humans compared with mice would engender these age and/or species variances (for review, see ref. 42). Moreover, these mechanisms do not readily explain one of the most distinctive

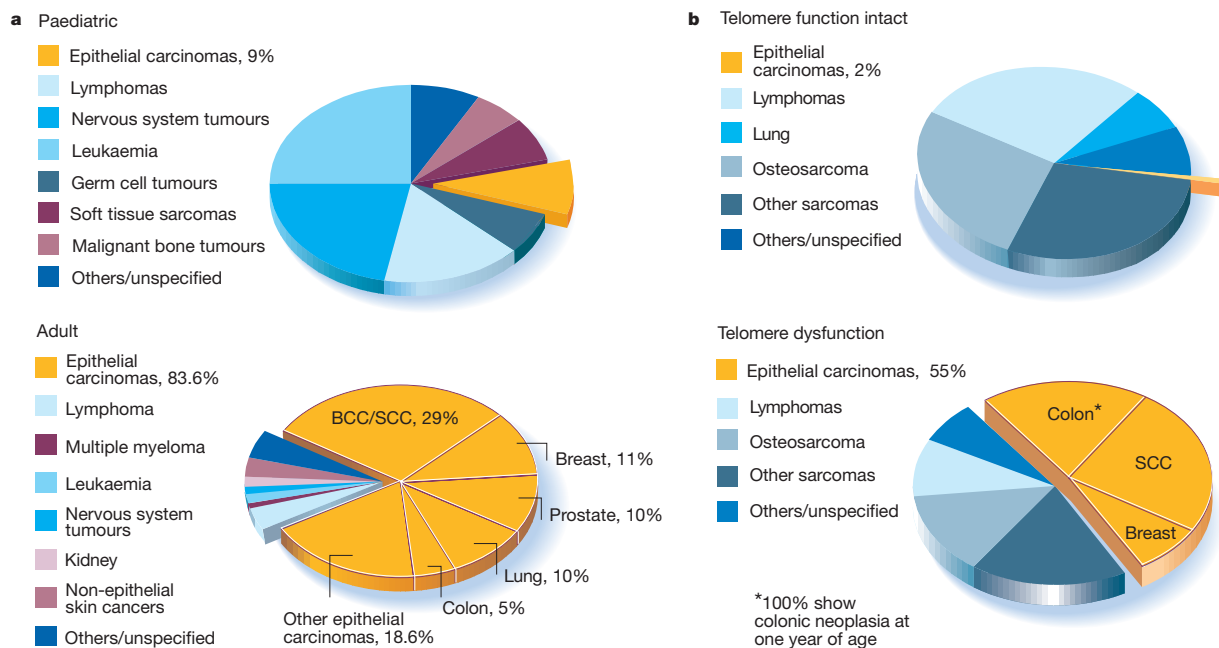


Figure 2 Tumour spectrum in human and mouse. **a**, Tumour spectra in humans represented in pie charts. Top chart includes the paediatric population; bottom chart the adult (male + female) population. Note the marked increase in proportion of epithelial carcinomas (orange-shaded segments) in the adult population. Compiled based on data in ref. 1. For basal cell carcinomas (BCC) and squamous cell carcinomas (SCC), an estimated total of 500,000 cases is used⁴⁵. **b**, Clinically apparent tumour spectra in mouse with and without functional telomeres, represented in pie charts⁴⁴. Epithelial carcinomas are included in the orange-shaded segments.

features of epithelial carcinomas in ageing humans, as opposed to paediatric cancers and virtually all mouse cancers (including occasional epithelial carcinomas) — a radically abnormal cytogenetic profile typified by marked aneuploidy and complex non-reciprocal translocations⁴³.

A recent study using the telomerase-knockout mouse has indicated that differences in telomere length and regulation might impact dramatically on both the spectrum and cytogenetics of tumours during ageing⁴⁴. Specifically, ageing telomerase-deficient mice, heterozygous for mutant p53, exhibited a pronounced shift in their tumour spectrum to carcinomas of the breast, colon and skin (Fig. 2). Importantly, these cancers emerged with cytogenetic profiles typical of human carcinomas. Are these data from the telomerase-deficient mouse relevant to mechanisms of epithelial carcinogenesis in aged humans? The answer has yet to be determined but recent cytogenetic and telomerase data from staged human epithelial cancers, particularly breast and colon cancers, makes the telomere–carcinoma connection an intriguing concept.

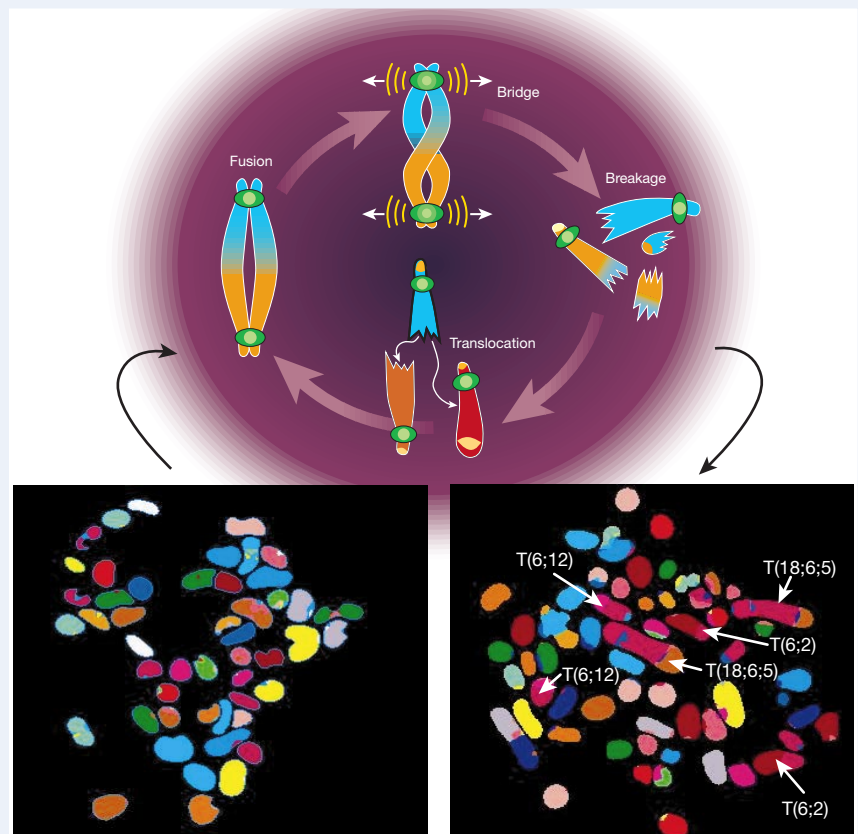
Telomeres comprise the nucleoprotein complexes that cap the ends of eukaryotic chromosomes and are maintained by the specialized reverse transcriptase, telomerase⁴⁵. In human cells, insufficient levels of telomerase lead to gradual telomere attrition with passage in culture⁴⁶ and possibly with ageing and tumorigenesis *in vivo*^{47–49}. Telomeres in human cancer cells are often significantly shorter than their normal tissue counterparts, indicating that telomere attrition occurs during the formative stages of cancer development where cell division takes place in the setting of low or absent telomerase activity⁴⁷. The subsequent reactivation of telomerase would restore telomere function, albeit at a shorter set length. In the mouse, significant telomere shortening and dysfunction does not take place owing to somatic telomerase expression and long telomere lengths^{50–54}. The observation of a more ‘humanized’ tumour spectrum and cytogenetics in mice engineered to experience age-dependent telomere

attrition has led to the speculation that telomere dysfunction might be one of several mechanisms driving epithelial carcinogenesis as humans advance in age (Fig. 2).

Epithelial carcinomas derive from cell lineages that undergo continual renewal throughout life. Against this backdrop of presumed age-dependent telomere attrition, somatic mutations driving aberrant epithelial proliferation would lead to an overall shorter telomere length. If somatic mutations also neutralize the retinoblastoma/INK4a/p53-dependent senescent checkpoint, then continued growth beyond the Hayflick limit (the life span of normal fibroblasts *in vitro*) coupled with progressive telomere erosion would culminate in ‘cellular crisis’ — a period of severe telomere dysfunction accompanied by rampant genomic instability and massive cell death⁵⁵. This dysfunctional telomere-induced genomic instability (or telomeric instability), combined with reactivation of telomerase, might enable a subset of post-crisis cells to emerge with a genetic profile permissive for malignant progression. In this manner, telomere-based crisis could provide the means to rapidly generate mutations necessary to initiate carcinogenesis. Subsequent reactivation of telomerase would serve to quell severe chromosomal instability and generate a more stable genome wherein the accumulation of mutational events further drives advanced stages of tumour progression. Such a transient period of explosive chromosomal instability could help generate the relatively high number of mutations thought to be required for adult epithelial carcinogenesis^{14,15}. Moreover, telomere dysfunction and associated formation of dicentric chromosomes would set in motion fusion–bridge–breakage cycles⁵⁶, a mechanism capable of producing rapid and widespread changes in gene dosage as well as complex cytogenetics⁴⁴ (Fig. 3).

This telomeric instability model of epithelial carcinogenesis fits well with what is known about the timing of telomerase activation and evolving genomic changes during various stages of human carcinoma development, particularly those of the breast and

Figure 3 Fusion–bridge–breakage mechanism and cytogenetic profiles. Fusion–bridge–breakage process leads to chromosomal fragmentation and non-reciprocal translocations (top). Spectral karyotype (SKY) profile of mouse tumour cells with functional telomeres on left and with dysfunctional telomeres on right.



colon (Fig. 4). Comparative genome hybridization (CGH) has shown that human breast and colon tumours sustain widespread gains and losses of regions of chromosomes early in their development—such ploidy changes detected by CGH correlate well with the presence of complex chromosomal rearrangements^{57–59}. This phase of genomic instability is evident by the carcinoma-*in-situ* stage of breast cancer and the early adenoma stage of colorectal cancer²⁰. As these cancers progress through invasive and metastatic stages, genomic instability continues, apparently at a moderate rate^{57–59}, but is predicted to occur through non-telomere-based mechanisms. Correspondingly, the measurement of telomerase activity in adenomatous polyps and colorectal cancers has established that telomerase activity is low or undetectable in small and intermediate sized polyps and increases markedly in large adenomas and colorectal carcinomas⁶⁰. Therefore, it seems that there is widespread and severe chromosomal instability early on during human tumorigenesis at a time when telomerase activity is low. However, it is worth emphasizing that a role for telomere dysfunction in human epithelial carcinogenesis will require an analysis of chromosomal/telomere status and telomerase activity levels in the same tumour samples from humans. Such studies are needed to fortify a causal link between telomere dysfunction and early chromosomal instability in human neoplasms.

Why should the pro-tumorigenic aspects of telomere dysfunction exert their action most prominently in self-renewing epithelial compartments and not contribute to an equivalent increase in other highly proliferative lineages such as lymphoid cells or mesenchymal lineages? One possible factor may relate to the extreme sensitivity of lymphoid cells to even modest levels of telomere dysfunction—impaired mitogenic responsiveness and increased apoptosis of lymphocytes is observed in young fourth-generation telomerase RNA component (mTERC)-null mice⁶¹. In contrast, compromise of gastrointestinal epithelium, as evidenced by intestinal villus atrophy,

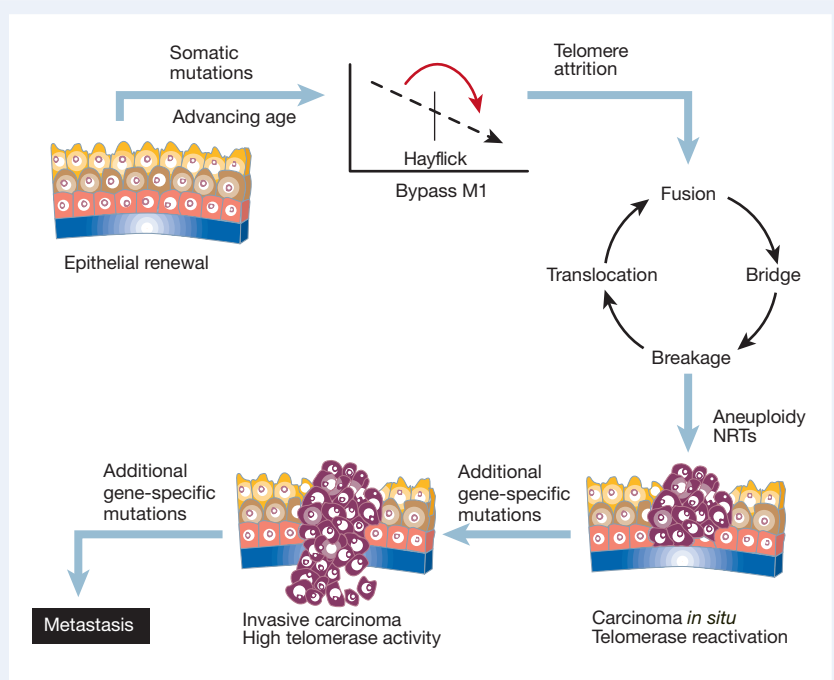
becomes apparent only in aged sixth-generation mTERC-null mice⁶². It is intriguing that telomerase is more readily upregulated following stimulation of normal lymphocytes⁶³ and that Myc, a potent regulator of telomerase, is invariably dysregulated in lymphoid malignancies⁶⁴. Thus, the ready activation of telomerase and consistent Myc dysregulation could avoid telomere-based crisis during lymphomagenesis and might also explain their less complex cytogenetic profiles of paediatric lymphoid malignancies. With regard to mesenchymal tumours, the lack of a pronounced increase could presumably relate to the relatively low proliferative rate of these cell types, leading to reduced levels of age-dependent telomere attrition as well as a reduced expansion of stem/progenitor cells bearing a mutational load.

Finally, this model of telomere-related instability provides a rational explanation for the high incidence of cancer associated with diseases characterized by chronic cell destruction and renewal. One of the most notable examples of this strong link is the high incidence of hepatocellular carcinoma in late stage cirrhotic livers in which decades of continual hepatocyte destruction and regeneration are associated with a reduction of telomeres to a critically short length^{65–68}. Such observations have raised an unanticipated therapeutic opportunity in that early somatic reconstitution of telomerase could attenuate telomere attrition and paradoxically reduce the occurrence of cancers in high-turnover disease states⁶⁵, a supposition that will require additional preclinical studies.

Tumour biology in the ageing organism

Malignant transformation represents the phenotypic end-point of successive genetic lesions impacting on functionally diverse cancer-relevant pathways. For virtually all cancer cells, there seem to be certain rites of passage on the path to transformation that include loss of cell-cycle control as a result of disruption of the retinoblastoma pathway⁶⁹, neutralization of a p53-dependent apoptotic

Figure 4 Dysfunctional telomere-induced genomic instability model of epithelial carcinogenesis. Continuous epithelial turnover during ageing is thought to lead to telomere shortening. When coupled with somatic mutations inactivating retinoblastoma/INK4a/p53 checkpoints, the Hayflick limit (mortality stage 1 (M1) or replicative senescence) can be bypassed. Continuous proliferation beyond the Hayflick limit results in progressive telomere attrition and subsequent fusion–bridge–breakage cycles in cells with dysfunctional telomeres. This process culminates in aneuploidy and complex non-reciprocal translocations (NRTs), resulting in massive and rapid changes in gene dosage observed in early carcinogenesis. Telomerase reactivation in the carcinoma-*in-situ* stage leads to relative chromosomal stability, providing a genome in which additional gene-specific mutations arise that are critical for progression to invasive and metastatic stages.



response elicited by this aberrant proliferation^{70–72}, long-term maintenance of chromosomal integrity^{73–76}, and establishment of a pro-tumorigenic microenvironment (reviewed in refs 16, 77). The latter transition is less well understood but has become increasingly recognized as a critical aspect of the tumorigenic process.

Complex homotypic and heterotypic interactions take place between tumour and host cells. The fully malignant tumour cell must become competent at instructing the host to establish a permissive and supportive environment for tumour growth and maintenance, including instructions to grow new blood vessels or to ignore tumour-associated antigens. The concept that genetic lesions intrinsic to the cancer cell drive and sustain these pro-tumorigenic host responses has received direct experimental support from inducible mouse models of cancer in which tumour-derived oncogenic signals were shown to be continually required for survival of host-derived endothelial cells^{16,78,79}. Pro-tumorigenic signals emanating from the normal host compartment are less well characterized, nor is it clear whether this compartment evolves, with advancing age, towards one that is more permissive for carcinogenesis. One recurring example of a host-derived change might be a decline in immune function, one that would allow for increased escape of tumour cells from immune surveillance. This notion is consistent with a lower incidence of lymphomas in ageing mice exhibiting increased immune vigour⁸⁰, although these linkage studies do not rule out the possibility of having a tumour modifier linked tightly to the major histocompatibility complex locus. Is there evidence for the existence of other cancer-relevant age-dependent host factors, beyond possible immune deterioration? Some answers may lie in the complex interactions among stroma, epithelium and inflammatory cells.

Stromal–epithelial interactions are known to be crucial in the growth, survival and differentiation of normal and neoplastic epithelial cells. Fibroblasts are a major constituent of the stromal compartment and exhibit a number of age-dependent changes (see below). The extent to which these changes influence the stromal environment and affect the threshold for epithelial carcinogenesis is not yet known. However, an integration of several observations from the fields of ageing, development and tumour biology raises some intriguing possibilities. First comes the evidence that stromal cells exert a profound effect on tumorigenesis, particularly epithelial carcinoma⁸¹. In an elegant series of reconstitution experiments, Cunha and colleagues assessed the explant tumour growth derived

from fully transformed prostatic epithelial cells admixed with fibroblasts isolated from either prostate cancer stroma or normal prostatic stroma⁸². Only prostatic carcinoma-associated fibroblasts were shown to support the tumorigenic growth of fully transformed prostatic epithelial cells *in vivo*⁸², indicating that non-transformed stromal cells derived from the tumour microenvironment have either lost the capacity to exert suppressive control over fully initiated epithelial cells and/or acquired new capabilities permissive for tumorigenesis.

The signals responsible for ‘reprogramming’ of non-transformed stromal fibroblasts are not known and presumably emanate from tumour cells, although another, not mutually exclusive, possibility is that age-dependent processes *per se* entrain stromal cells. The latter notion follows from the observed accumulation of senescent dermal fibroblasts over time⁸³ and their increased production of cytokines, proteases and other matrix-degrading enzymes^{27,84,85}. The factors driving fibroblast senescence *in vivo* have not been identified but could be analogous to those evoking senescence in cultured cells such as mitotic misregulation²⁷, aberrant oncogenic signals⁸⁶, and/or intrinsic DNA damage⁸⁷.

Regardless of the signals involved, the accumulation of senescent stromal cells with ageing, coupled with altered expression and elaboration of proteases and cytokines, has been hypothesized to create a tissue microenvironment that is more permissive for the growth of oncogenically transformed epithelial cells⁸⁵. This hypothesis, proposed first by Campisi, seems more attractive in the light of a series of landmark studies of tumour biology in two transgenic mouse models of cancer^{88–90}. In these studies, Hanahan and colleagues have shown that the matrix metalloprotease, MMP-9/Gel-B, acts as a key paracrine regulator of carcinogenesis. Furthermore, the effect of MMP-9 was reconstituted by bone-marrow transplantation onto an MMP-9 gene-knockout background, providing evidence for the sufficiency of MMP-9 and the fact that inflammatory cells serve as the main source of MMP-9. These model systems have also revealed that MMP-9 acts by mobilizing a latent store of vascular endothelial growth factor, the most potent endothelial cell mitogen and survival factor, thereby promoting the angiogenic switch critical for tumour progression⁸⁹. It is worth noting that, although expression of MMP-9 has not been examined, several enzymes involved in extracellular matrix remodelling (for example, interstitial collagenase and stromelysin) are overexpressed in senescent fibroblasts^{27,84,91}.

Moreover, overexpression of MMP-3/stromelysin-1 by the stroma has been shown to promote mammary carcinogenesis⁹². These experimental observations, combined with 'stromal senescence', raise the question of whether age-dependent changes in the stromal milieu could contribute to tumorigenesis, specifically angiogenesis, in part by modulating inflammatory cell recruitment and behaviour (through elaboration of stroma-derived cytokines) or by acting in concert with inflammatory cells through increased local production of protease.

Summary, challenges and opportunities

Here I have reviewed some of the issues that are relevant to the connection between ageing and cancer and have speculated on how various age-sensitive molecular and physiological processes could act in concert to promote cancer, specifically epithelial carcinogenesis. Current data are consistent with, but not proof for, the combined pathogenetic effects of mutation load, epigenetic regulation, telomeric instability and altered stromal milieu. When considered collectively, these processes form a scenario in which continual epithelial renewal and somatic mutations (which disable the senescence checkpoint) allow for unrestrained growth and telomere attrition, culminating in a fusion-bridge-breakage-translocation process and marked genetic alterations. The relative resistance of epithelial cells to early telomere crisis, combined with the frequent loss of p53 in epithelial neoplasms, would serve to expand the pool of crisis cells from which a pro-cancer genotype would emerge. After reactivation of telomerase, this initiated cancer cell population incurs additional mutations essential for progression towards full transformation. The capacity of malignantly transformed epithelium to generate a mature tumour would also be governed by the tumour cell's ability to overcome the suppressive stromal milieu, a barrier that could be breached more readily in cytokine/protease-rich senescent stroma.

Each of these mechanisms, from molecular to cellular to organismal, require further verification through rigorous experimentation but their pursuit should uncover new paradigms from which rational therapeutics will emerge. In each of the areas presented in this perspective, there exist critical issues that need to be addressed. Mutation-rate analysis will benefit from improved *in vivo* methods that will determine accurately and comprehensively the frequency and spectrum of mutations in normal human cells and tissues, particularly in tissue stem cells. The development of equivalent assays in the mouse should provide comparative cross-species information that will contribute to our understanding of the basis for the distinct tumour spectrum seen in ageing humans and mice. In the area of epigenetics, unmet challenges include the needs for a detailed understanding of the methylation machinery and how its dysregulation contributes to carcinogenesis, a more comprehensive genome-wide view of the genes and loci subject to CpG-island methylation in cancer cells, and a determination of whether and how DNA methylation influences genome stability⁹³. The role of telomere dysfunction in human epithelial carcinogenesis remains unproven. As such, it will be important to document telomere attrition in renewing epithelial stem cells and to perform a simultaneous comparison of telomere status, telomerase activity and chromosomal instability, particularly during the earliest stages of human epithelial cell transformation. Finally, the hypothesis that senescent stroma might facilitate epithelial carcinogenesis should be testable in reconstitution studies that assess the malignant potential of initiated epithelial cells admixed with either senescent or young stromal cells.

Cancer extracts a significant social and economic toll. The American Cancer Society estimates that, in the United States, 1,220,100 new cancer cases will be diagnosed in the year 2000 and nearly 80% of all cancers will arise in individuals aged 55 and older¹. Each day more than 1,500 individuals succumb to this disease¹, most after a protracted and expensive clinical course. As the number of older individuals increases steadily in the population, the development of effective chemopreventive measures directed squarely

against epithelial carcinogenesis assumes increasing significance. Productive clinical advances will no doubt flow from an improved understanding of the molecular and cellular mechanisms driving cancer in aged epithelium. A view of such mechanisms is coming into focus but many important details have yet to be revealed. It is with high confidence that the cancer research community anticipates that mechanistic insights will in turn lead to the design of effective therapeutics for prevention and for preferential tumour cell death — therapeutics that promise to convert this age of cancer to one concerned solely with longevity. □

1. American Cancer Society. *Cancer Facts and Figures 2000* 1–7 (American Cancer Society, Atlanta, 2000).
2. Bishop, J. M. Molecular themes in oncogenesis. *Cell* **64**, 235–248 (1991).
3. Nowell, P. C. The clonal evolution of tumor cell populations. *Science* **194**, 23–28 (1976).
4. Vogelstein, B. *et al.* Genetic alterations during colorectal tumor development. *N. Engl. J. Med.* **319**, 525–532 (1988).
5. Balaban, G. B., Herlyn, M., Clark, W. H. Jr & Nowell, P. C. Karyotypic evolution in human malignant melanoma. *Cancer Genet. Cytogenet.* **19**, 113–122 (1986).
6. Naylor, S. L., Johnson, B. E., Minna, J. D. & Sakaguchi, A. Y. Loss of heterozygosity of chromosome 3p markers in small-cell lung cancer. *Nature* **329**, 451–454 (1987).
7. James, C. D. *et al.* Clonal genomic alterations in glioma malignancy stages. *Cancer Res.* **48**, 5546–5551 (1988).
8. Rabinovitch, P. S., Reid, B. J., Haggitt, R. C., Norwood, T. H. & Rubin, C. E. Progression to cancer in Barrett's esophagus is associated with genomic instability. *Lab. Invest.* **60**, 65–71 (1988).
9. Harman, D. Aging: a theory based on free-radical and radiation chemistry. *J. Gerontol.* **6**, 298–300 (1956).
10. Ames, B. N., Shigenaga, M. K. & Hagen, T. M. Oxidants, antioxidants and the degenerative diseases of aging. *Proc. Natl Acad. Sci. USA* **90**, 7915–7922 (1993).
11. Minnick, D. T. & Kunkel, T. A. DNA synthesis errors, mutators and cancer. *Cancer Surv.* **28**, 3–20 (1996).
12. Oller, A. R., Rastogi, P., Morgenthaler, S. & Thilly, W. G. A statistical model to estimate variance in long term-low dose mutation assays: testing of the model in a human lymphoblastoid mutation assay. *Mutat. Res.* **216**, 149–161 (1989).
13. Jackson, A. L. & Loeb, L. A. The mutation rate and cancer. *Genetics* **148**, 1483–1490 (1998).
14. Armitage, P. & Doll, R. The age distribution of cancer and a multi-stage theory of carcinogenesis. *Br. J. Cancer* **8**, 1–12 (1954).
15. Renan, M. J. How many mutations are required for tumorigenesis? Implications from human cancer data. *Mol. Carcinog.* **7**, 139–146 (1993).
16. Hanahan, D. & Weinberg, R. A. The hallmarks of cancer. *Cell* **100**, 57–70 (2000).
17. Loeb, L. A. Mutator phenotype may be required for multistage carcinogenesis. *Cancer Res.* **51**, 3075–3079 (1991).
18. Kinzler, K. W. & Vogelstein, B. Lessons from hereditary colorectal cancer. *Cell* **87**, 159–170 (1996).
19. Lengauer, C., Kinzler, K. W. & Vogelstein, B. Genetic instability in colorectal cancers. *Nature* **386**, 623–627 (1997).
20. Gray, J. W. & Collins, C. Genome changes and gene expression in human solid tumors. *Carcinogenesis* **21**, 443–452 (2000).
21. Weinberg, R. The cat and mouse games that genes, viruses, and cells play. *Cell* **88**, 573–575 (1997).
22. Turker, M. S. Estimation of mutation frequencies in normal mammalian cells and the development of cancer. *Semin. Cancer Biol.* **8**, 407–419 (1998).
23. Dolle, M. E. *et al.* Rapid accumulation of genome rearrangements in liver but not in brain of old mice. *Nature Genet.* **17**, 431–434 (1997).
24. Dolle, M. E., Snyder, W. K., Gossen, J. A., Lohman, P. H. & Vijg, J. Distinct spectra of somatic mutations accumulated with age in mouse heart and small intestine. *Proc. Natl Acad. Sci. USA* **97**, 8403–8408 (2000).
25. Vijg, J. & Dolle, M. E. T. in *Handbook of the Biology of Ageing* 5th edn (eds Masoro, E. J. & Austad, S. N.) (Academic, in the press).
26. Curtis, H. & Crowley, C. Chromosome aberrations in liver cells in relation to the somatic mutation theory of aging. *Radiat. Res.* **19**, 337–344 (1963).
27. Ly, D. H., Lockhart, D. J., Lerner, R. A. & Schultz, P. G. Mitotic misregulation and human aging. *Science* **287**, 2486–2492 (2000).
28. Ramsey, M. J. *et al.* The effects of age and lifestyle factors on the accumulation of cytogenetic damage as measured by chromosome painting. *Mutat. Res.* **338**, 95–106 (1995).
29. Jacobs, P., Brunton, M., Brown, W., Doll, R. & Goldstein, H. Change of human chromosome count distributions with age: evidence for a sex difference. *Nature* **197**, 1080–1081 (1963).
30. Tucker, J. D., Spruill, M. D., Ramsey, M. J., Director, A. D. & Nath, J. Frequency of spontaneous chromosome aberrations in mice: effects of age. *Mutat. Res.* **425**, 135–141 (1999).
31. Moriawaki, S., Ray, S., Tarone, R. E., Kraemer, K. H. & Grossman, L. The effect of donor age on the processing of UV-damaged DNA by cultured human cells: reduced DNA repair capacity and increased DNA mutability. *Mutat. Res.* **364**, 117–123 (1996).
32. Goukassian, D. *et al.* Mechanisms and implications of the age-associated decrease in DNA repair capacity. *FASEB J.* **14**, 1325–1334 (2000).
33. Liu, S. C., Parsons, C. S. & Hanawalt, P. C. DNA repair response in human epidermal keratinocytes from donors of different age. *J. Invest. Dermatol.* **79**, 330–335 (1982).
34. Kolodner, R. D. *et al.* Germ-line msh6 mutations in colorectal cancer families. *Cancer Res.* **59**, 5068–5074 (1999).
35. Baylin, S. B. & Herman, J. G. DNA hypermethylation in tumorigenesis: epigenetics joins genetics. *Trends Genet.* **16**, 168–174 (2000).
36. Jones, P. A. & Laird, P. W. Cancer epigenetics comes of age. *Nature Genet.* **21**, 163–167 (1999).
37. Issa, J. P. *et al.* Methylation of the oestrogen receptor CpG island links ageing and neoplasia in human colon. *Nature Genet.* **7**, 536–540 (1994).
38. Issa, J. P. Aging, DNA methylation and cancer. *Crit. Rev. Oncol. Hematol.* **32**, 31–43 (1999).
39. Cormier, R. T. & Dove, W. F. Dnmt1N/+ reduces the net growth rate and multiplicity of intestinal adenomas in C57BL/6-multiple intestinal neoplasia (Min)/+ mice independently of p53 but demonstrates strong synergy with the modifier of Min 1 (AKR) resistance allele. *Cancer Res.* **60**, 3965–3970 (2000).

40. Laird, P. W. *et al.* Suppression of intestinal neoplasia by DNA hypomethylation. *Cell* **81**, 197–205 (1995).
41. Rhee, I. *et al.* CpG methylation is maintained in human cancer cells lacking DNMT1. *Nature* **404**, 1003–1007 (2000).
42. DePinho, R. A. & Sherr, C. J. Cellular senescence: mitotic clock or culture shock? *Cell* **102**, 407–410 (2000).
43. Atkin, N. B. Solid tumor cytogenetics. Progress since 1979. *Cancer Genet. Cytogenet.* **40**, 3–12 (1989).
44. Artandi, S. E. *et al.* Telomere dysfunction promotes non-reciprocal translocations and epithelial cancers in mice. *Nature* **406**, 641–645 (2000).
45. Greider, C. Telomerase activity, cell proliferation, and cancer. *Proc. Natl Acad. Sci. USA* **95**, 90 (1998).
46. Harley, C. B., Futcher, A. B. & Greider, C. W. Telomeres shorten during ageing of human fibroblasts. *Nature* **345**, 458–460 (1990).
47. Hastie, N. D. *et al.* Telomere reduction in human colorectal carcinoma and with ageing. *Nature* **346**, 866–868 (1990).
48. Allsopp, R. C. *et al.* Telomere length predicts replicative capacity of human fibroblasts. *Proc. Natl Acad. Sci. USA* **89**, 10114–10118 (1992).
49. Chadeneau, C., Hay, K., Hirte, H. W., Gallinger, S. & Bacchetti, S. Telomerase activity associated with acquisition of malignancy in human colorectal cancer. *Cancer Res.* **55**, 2533–2536 (1995).
50. Kipling, D. & Cooke, H. J. Hypervariable ultra-long telomeres in mice. *Nature* **347**, 400–402 (1990).
51. Prowse, K. R. & Greider, C. W. Developmental and tissue-specific regulation of mouse telomerase and telomere length. *Proc. Natl Acad. Sci. USA* **92**, 4818–4822 (1995).
52. Broccoli, D., Godley, L. A., Donehower, L. A., Varmus, H. E. & de, L. T. Telomerase activation in mouse mammary tumors: lack of detectable telomere shortening and evidence for regulation of telomerase RNA with cell proliferation. *Mol. Cell. Biol.* **16**, 3765–3772 (1996).
53. Greenberg, R. A., Allsopp, R. C., Chin, L., Morin, G. B. & DePinho, R. A. Expression of mouse telomerase reverse transcriptase during development, differentiation and proliferation. *Oncogene* **16**, 1723–1730 (1998).
54. Martin-Rivera, L., Herrera, E., Albar, J. & Blasco, M. A. Expression of mouse telomerase catalytic subunit in embryos and adult tissues. *Proc. Natl Acad. Sci. USA* **95**, 10471–10476 (1998).
55. Counter, C. M. *et al.* Telomere shortening associated with chromosome instability is arrested in immortal cells which express telomerase activity. *EMBO J.* **11**, 1921–1929 (1992).
56. McClintock, B. The stability of broken ends of chromosomes in *Zea mays*. *Genetics* **26**, 234–282 (1941).
57. Nishizaki, T. *et al.* Genetic alterations in primary breast cancers and their metastases: direct comparison using modified comparative genomic hybridization. *Genes Chromosomes Cancer* **19**, 267–272 (1997).
58. Buerger, H. *et al.* Comparative genomic hybridization of ductal carcinoma in situ of the breast: evidence of multiple genetic pathways. *J. Pathol.* **187**, 396–402 (1999).
59. Al-Mulla, F., Keith, W. N., Pickford, I. R., Going, J. J. & Birnie, G. D. Comparative genomic hybridization analysis of primary colorectal carcinomas and their synchronous metastases. *Genes Chromosomes Cancer* **24**, 306–314 (1999).
60. Tang, R., Cheng, A. J., Wang, J. Y. & Wang, T. C. Close correlation between telomerase expression and adenomatous polyp progression in multistep colorectal carcinogenesis. *Cancer Res.* **58**, 4052–4054 (1998).
61. Lee, H. W. *et al.* Essential role of mouse telomerase in highly proliferative organs. *Nature* **392**, 569–574 (1998).
62. Rudolph, K. L. *et al.* Longevity, stress response, and cancer in aging telomerase-deficient mice. *Cell* **96**, 701–712 (1999).
63. de Lange, T. Activation of telomerase in a human tumor. *Proc. Natl Acad. Sci. USA* **91**, 2882–2885 (1994).
64. Foley, K. P. & Eisenman, R. N. Two MAD tails: what the recent knockouts of Mad1 and Mx1 tell us about the MYC/MAX/MAD network. *Biochim. Biophys. Acta* **1423**, M37–M47 (1999).
65. Rudolph, K. L., Chang, S., Millard, M., Schreiber-Agus, N. & DePinho, R. A. Inhibition of experimental liver cirrhosis in mice by telomerase gene delivery. *Science* **287**, 1253–1258 (2000).
66. Kitada, T., Seki, S., Kawakita, N., Kuroki, T. & Monna, T. Telomere shortening in chronic liver diseases. *Biochem. Biophys. Res. Commun.* **211**, 33–39 (1995).
67. Miura, N. *et al.* Progressive telomere shortening and telomerase reactivation during hepatocellular carcinogenesis. *Cancer Genet. Cytogenet.* **93**, 56–62 (1997).
68. Urabe, Y. *et al.* Telomere length in human liver diseases. *Liver* **16**, 293–297 (1996).
69. Weinberg, R. A. The retinoblastoma protein and cell cycle control. *Cell* **81**, 323–330 (1995).
70. Debbas, M. & White, E. Wild-type p53 mediates apoptosis by E1A, which is inhibited by E1B. *Genes Dev.* **7**, 546–554 (1993).
71. Symonds, H. *et al.* p53-dependent apoptosis suppresses tumor growth and progression in vivo. *Cell* **78**, 703–711 (1994).
72. Morgenbesser, S. D., Williams, B. O., Jacks, T. & DePinho, R. A. p53-dependent apoptosis produced by Rb-deficiency in the developing mouse lens. *Nature* **371**, 72–74 (1994).
73. Kim, N. W. *et al.* Specific association of human telomerase activity with immortal cells and cancer. *Science* **266**, 2011–2015 (1994).
74. Blasco, M. A., Rizen, M., Greider, C. W. & Hanahan, D. Differential regulation of telomerase activity and telomerase RNA during multi-stage tumorigenesis. *Nature Genet.* **12**, 200–204 (1996).
75. Greenberg, R. A. *et al.* Short dysfunctional telomeres impair tumorigenesis in the INK4a(Δ 2/3) cancer-prone mouse. *Cell* **97**, 515–525 (1999).
76. Hahn, W. C. *et al.* Creation of human tumour cells with defined genetic elements. *Nature* **400**, 464–468 (1999).
77. Chin, L. & DePinho, R. A. Flipping the oncogene switch: illumination of tumor maintenance and regression. *Trends Genet.* **16**, 147–150 (2000).
78. Chin, L. *et al.* Essential role for oncogenic RAS in tumour maintenance. *Nature* **400**, 468–472 (1999).
79. Pelengaris, S., Littlewood, T., Khan, M., Elia, G. & Evan, G. Reversible activation of c-Myc in skin: induction of a complex neoplastic phenotype by a single oncogenic lesion. *Mol. Cell* **3**, 565–577 (1999).
80. Gelman, R., Watson, A., Bronson, R. & Yunis, E. Murine chromosomal regions correlated with longevity. *Genetics* **118**, 693–704 (1988).
81. Cunha, G. & Hom, Y. Role of mesenchymal-epithelial interactions in mammary gland development. *J. Mamm. Gland Biol. Neoplasia* **1**, 21–35 (1996).
82. Olumi, A. F. *et al.* Carcinoma-associated fibroblasts direct tumor progression of initiated human prostatic epithelium. *Cancer Res.* **59**, 5002–5011 (1999).
83. Dimri, G. P. *et al.* A biomarker that identifies senescent human cells in culture and in aging skin in vivo. *Proc. Natl Acad. Sci. USA* **92**, 9363–9367 (1995).
84. Shelton, D. N., Chang, E., Whittier, P. S., Choi, D. & Funk, W. D. Microarray analysis of replicative senescence. *Curr. Biol.* **9**, 939–945 (1999).
85. Campisi, J. Aging and cancer: the double-edged sword of replicative senescence. *J. Am. Geriatr. Soc.* **45**, 482–488 (1997).
86. Serrano, M., Lin, A. W., McCurrach, M. E., Beach, D. & Lowe, S. W. Oncogenic ras provokes premature cell senescence associated with accumulation of p53 and p16INK4a. *Cell* **88**, 593–602 (1997).
87. Di Leonardo, A., Linke, S. P., Clarkin, K. & Wahl, G. M. DNA damage triggers a prolonged p53-dependent G1 arrest and long-term induction of Cip1 in normal human fibroblasts. *Genes Dev.* **8**, 2540–2551 (1994).
88. Coussens, L. M. *et al.* Inflammatory mast cells up-regulate angiogenesis during squamous epithelial carcinogenesis. *Genes Dev.* **13**, 1382–1397 (1999).
89. Bergers, G. *et al.* Matrix metalloproteinase-9 triggers the angiogenic switch during carcinogenesis. *Nature Cell Biol.* **2**, 737–744 (2000).
90. Coussens, L., Tinkle, C., Hanahan, D. & Werb, Z. MMP-9 supplied by bone marrow-derived cells contributes to skin carcinogenesis. *Cell* (in the press).
91. Millis, A. J., Hoyle, M., McCue, H. M. & Martini, H. Differential expression of metalloproteinase and tissue inhibitor of metalloproteinase genes in aged human fibroblasts. *Exp. Cell Res.* **201**, 373–379 (1992).
92. Sternlicht, M. D. *et al.* The stromal proteinase MMP3/stromelysin-1 promotes mammary carcinogenesis. *Cell* **98**, 137–146 (1999).
93. Robertson, K. D. & Jones, P. A. DNA methylation: past, present and future directions. *Carcinogenesis* **21**, 461–467 (2000).
94. Piantanelli, L. Cancer and aging: from the kinetics of biological parameters to the kinetics of cancer incidence and mortality. *Ann. NY Acad. Sci.* **521**, 99–109 (1988).
95. Schwartz, R. A. *Skin Cancer: Recognition and Management* (Springer, New York, 1988).

Acknowledgements

I apologize to my colleagues whose relevant work I was unable to cite owing to space and reference limitations. I thank J. Vijg, D. Hanahan, J. Campisi, J.-P. Issa, N. Schreiber-Agus, G. Merlino, S. Mellis, L. Chin, S. Weiler and members of my laboratory for helpful discussions and critical comments. R.A.D. is supported by the National Institutes of Health and is an American Cancer Society Research Professor and a Steven and Michele Kirsch Foundation Investigator.

Genetic pathways that regulate ageing in model organisms

Leonard Guarente* & Cynthia Kenyon†

*Department of Biology, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA

†Department of Biochemistry and Biophysics, University of California at San Francisco, San Francisco, California 94143-0448, USA

Searches for genes involved in the ageing process have been made in genetically tractable model organisms such as yeast, the nematode *Caenorhabditis elegans*, *Drosophila melanogaster* fruitflies and mice. These genetic studies have established that ageing is indeed regulated by specific genes, and have allowed an analysis of the pathways involved, linking physiology, signal transduction and gene regulation. Intriguing similarities in the phenotypes of many of these mutants indicate that the mutations may also perturb regulatory systems that control ageing in higher organisms.

The propagation of a species could depend in principle either on the perfect maintenance of its individuals or on the ability to renew the population with young members. Nature most commonly has chosen the latter option. Renewal is typically achieved by the setting aside of a pristine lineage of genetic information — that is, the germ line — which is passed on by sexual reproduction. In contrast, the somatic lineage of all animals declines and degenerates with age, giving rise to phenotypic changes recognized as ageing. Studies in model organisms have begun to map out important genes and pathways that seem to regulate the pace of ageing and that are remarkably conserved from yeast to metazoans. These studies have the great advantage that one can use genetic approaches to search directly for mutations that change life span. This makes it possible to identify mechanisms in a way that is independent of any preconceived model of ageing. Recently, the analysis of such mutations that affect life span has revealed several different pathways that influence the ageing process. Our review will focus largely on insights into ageing that have been gleaned from analysing certain single-gene mutations in the two model organisms in which the molecular mechanisms of ageing are best understood, yeast and the nematode worm *C. elegans*. Studies reveal surprisingly simple and conserved mechanisms governing the pace of ageing and life span. Although it is too early to extrapolate these findings to mammals, it seems likely that they will spark molecular insights into causes of human ageing.

Caloric restriction

One of the best indications that the ageing process is subject to regulation is that life span can be extended by caloric restriction¹. Caloric restriction typically refers to a diet in which calories are limited by 30–40% compared with animals fed *ad libitum*. Caloric restriction extends life span in rodents, worms, yeast and probably primates. This response to caloric restriction has clear selective value, because it allows animals to postpone reproduction until food is available. However, when food is restored the animals can produce progeny, even when well-fed controls are post-reproductive or no longer alive. The mechanism by which caloric restriction extends life span is unclear. One hypothesis is that caloric restriction slows metabolism, thereby slowing the production of toxic reactive

oxygen species and, in turn, slowing ageing. Although this idea is consistent with a probable link between oxidative damage and ageing², it has never been validated experimentally. Other examples reinforce the notion that within a species there is a relationship between metabolic rate and ageing³. For example, *Drosophila* live much longer when maintained at lower temperatures. However, there is every reason to think that the correlation between metabolic rate and ageing can be broken by genetic changes. For example, bats have a comparable metabolic rate to mice, yet live more than ten times longer. In addition, single-gene mutations in yeast, *C. elegans*, *Drosophila* and mice can extend the life span significantly without an apparent slowing in metabolism. Thus it is reasonable to conclude that metabolic rate is important within an individual or species, but its impact may be overridden by evolutionary divergence (bats compared with mice) or even single-gene mutations conferring increased or greater longevity. It is interesting to speculate that mutations that break this correlation might alter the regulatory machinery that couples caloric restriction and ageing.

Gene silencing as a regulator of life span

Studies of ageing in the budding yeast *Saccharomyces cerevisiae* have led to the conclusion that a gene involved in the silencing of chromatin may be a key regulator of ageing. Silencing is a process by which entire regions of chromosomes encompassing blocks of genes are rendered transcriptionally inactive. In budding yeast the life span is measured by the number of cell divisions in a mother cell lineage^{4–6}, determined by microscopic manipulation (Fig. 1). Mother cells divide a relatively fixed number of times and undergo characteristic changes as they age. Although yeast seem to violate the rule of a demarcation of soma and germ line, closer inspection reveals a renewal process in the budding of daughter cells, from an ageing, soma-like lineage of mother cells. A phenotypic change that occurs in ageing mother cells is the expression of the repositories of α and α mating-type genes owing to a breakdown in genomic silencing at the *HML* and *HMR* loci where extra copies of this structural information is stored. Haploid cells normally express either a or α information from a mating-type locus termed *MAT*. The breakdown in silencing at *HML* and *HMR* leads to the simultaneous expression of genes of both mating types thereby resulting in sterility⁷. Other phenotypic changes evinced by mother cells as they

age include cellular enlargement and a slowing of the cell cycle. Gross changes in the morphology of the nuclei also occur, most notably a massive expansion in the size of the nucleolus (described in more detail below). Finally, bud scars, which are the remnant of each cellular division, accumulate on the surface of ageing mother cells.

Silencing in yeast is carried out by a complex of silent information regulator (Sir) proteins — Sir2, Sir3 and Sir4 — at telomeres and mating-type genes^{8,9}. A causal link between silencing and ageing was first made by the identification of the *SIR4-42* longevity mutation¹⁰, which redirects the Sir2/3/4 complex away from telomeres and *HM* loci to the region of the genome encoding the ribosomal RNA (rDNA). The rDNA contains 100–200 tandem repeats of genes encoding the large and small rRNAs and organizes the nucleolus as the site of ribosome synthesis. Sir2 normally mediates silencing in the rDNA without the other Sir proteins^{11,12}. Subsequent experiments validated the notion that the amount of Sir2 at the rDNA was predictive of life span; *sir2*-deletion strains have a short life span, and strains with an extra *SIR2* copy have super-extended life spans¹³. Thus an increase in rDNA silencing by Sir2 seems to increase life span. It has also been reported that deletion of the Rpd3 global histone deacetylase, which removes the acetyl groups on lysines of the amino termini of histones, extends life span¹⁴. Initially this finding seems strange, as the increase in acetylation of histones resulting from the loss of a deacetylase will normally result in an increase in gene expression. However, consistent with the extension in life span, the loss of Rpd3 actually leads to an increase in rDNA silencing¹⁵. The reason why removal of the Rpd3 deacetylase decreases expression is not clear, but may be an indirect effect of an increased expression of genes encoding the Sir silencing proteins.

Silenced chromatin in the repeated rDNA array represses recombination¹⁶, which would otherwise generate extrachromosomal rDNA circles (ERCs) that accumulate in subsequent divisions in mother cell lineages and have been reported to be a cause of ageing¹⁷. But the importance of ERCs in ageing has been a matter of some controversy. On the one hand, the generation of ERCs early by a site-specific recombinase shortens life span in mother cells¹⁷. Moreover, deletion of *FOB1*, encoding a protein that promotes a unidirectional block in rDNA replication, reduces or eliminates the formation of ERCs and extends life span¹⁸. These findings implicate ERCs in ageing in yeast. On the other hand, the short-lived *sgs1* mutant, defective in the Werners-like DNA helicase (see review in this issue by Martin and Oshima, pages 263–266), was recently reported not to have an elevated level of ERCs¹⁹. In addition, other DNA circles, such as the *TRP1 ARS1* plasmid, can shorten life span¹⁷, indicating that ERCs are not unique in their ability to accumulate in and kill mother cells. Finally, mother cells that experience a sojourn in stationary phase showed a shortened life span without any increase in ERCs, implicating at least one ERC-independent mechanism in ageing²⁰. The most accurate surmise at present is that ERCs are one of several inputs causing ageing in wild-type mother cells. The short life span in *sgs1* mutants, however, seems to be more complicated than originally surmised²¹, and may involve other mechanisms.

SNF1 is a candidate for a second input into ageing, acting through a metabolic pathway that determines carbon-source utilization and energy levels in cells. *SNF1* encodes a kinase that activates genes required for utilization of certain carbon sources other than glucose²². Surprisingly, the activity of *SNF1* was found to increase with the age of mother cells, even in the presence of abundant glucose²³. This change is accompanied by a metabolic shift to energy storage resulting in increases in cellular ATP and nicotinamide adenine dinucleotide (NAD). These changes are linked closely to the ageing process because null mutations in the *SNF1* co-repressor, *SIP2*, decrease life span, and mutations in the *SNF1* co-activator, *SNF4*, increase life span. It is possible that, when phosphorylated, one or more Snf1 targets cause accelerated ageing. The identification of such targets might suggest how this metabolic pathway is integrated with the silencing pathway above.

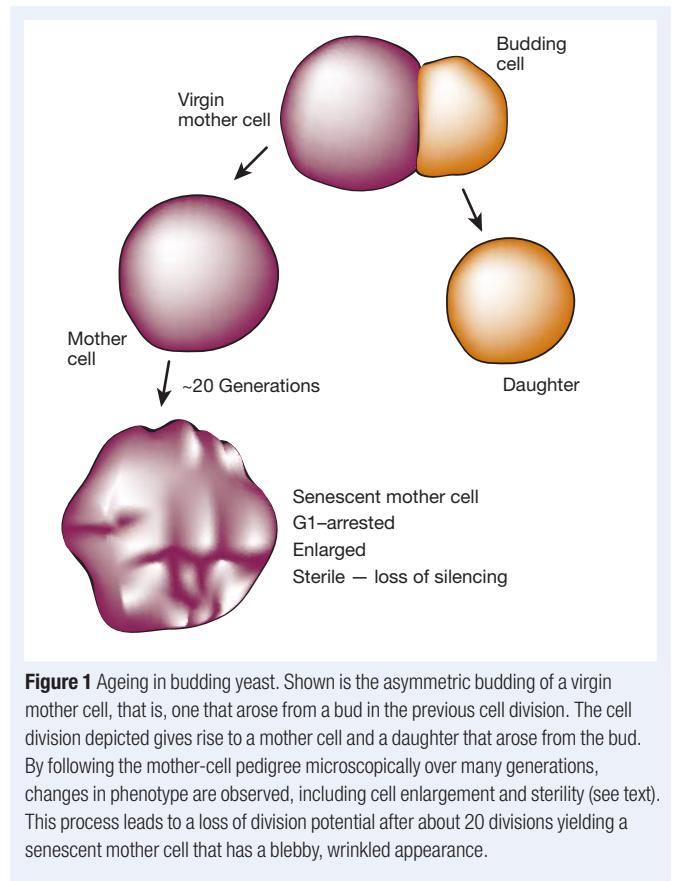


Figure 1 Ageing in budding yeast. Shown is the asymmetric budding of a virgin mother cell, that is, one that arose from a bud in the previous cell division. The cell division depicted gives rise to a mother cell and a daughter that arose from the bud. By following the mother-cell pedigree microscopically over many generations, changes in phenotype are observed, including cell enlargement and sterility (see text). This process leads to a loss of division potential after about 20 divisions yielding a senescent mother cell that has a blebby, wrinkled appearance.

Sir2 proteins at the interface of ageing and metabolism

The importance of silencing in ageing raises the question of what is the exact biochemical function of Sir2. A core domain of Sir2 has been widely conserved from bacteria to humans^{24,25} and possesses a fascinating enzymatic activity. The early observation that overexpression of Sir2 caused the global deacetylation of histones led to the conjecture that this protein was a histone deacetylase²⁶. However, biochemical experiments showed that the purified protein displayed a different enzymatic activity, albeit a weak one, that of an ADP ribosyltransferase, using NAD as a donor^{25,27}. An apparent solution to this puzzle came with the demonstration that Sir2 does indeed require NAD, but this requirement is for the activation of a robust histone deacetylase activity²⁸. This activity displayed specificity for lysines 9 and 14 of histone H3 and lysine 16 of H4, residues shown to be important for silencing *in vivo*. This NAD-dependent deacetylase activity seems to be a universal property of Sir2 proteins from bacteria to mammals^{28–30}.

The NAD requirement for the Sir2 deacetylase is highly unusual and provocative in the context of ageing. NAD is more typically used in hundreds of metabolic reactions in cells to receive electrons or, in its reduced form NADH, to donate them. In the case of Sir2, NAD seems to function as a cofactor that activates the deacetylase, potentially coupling that activity to the energy status of cells. Sir2 proteins may in this way sense metabolic rate and transduce this energy status to life span by their deacetylase activities. The increase in NAD levels in normal ageing mentioned above²³ might represent a homeostatic response of cells to help oppose the waning of genomic silencing by the Sir proteins.

In this regard, caloric restriction can be modelled in yeast by limiting glucose concentrations from 2% to 0.5% in the growth media or by mutating components of the cyclic AMP-dependent protein kinase A (PKA) pathway, which signals glucose availability to cells (Fig. 2)³¹. Either regimen leads to a significant extension in the life span of mother cells. Earlier studies implicated the Ras components

of this pathway in the determination of life span³². Limiting glucose in mutants of the PKA pathway did not elicit any further extension in life span, showing that these two regimens function in the same pathway. Interestingly, the extension in life span by caloric restriction in yeast was abolished by mutations in *SIR2* or *NPT1*, a gene involved in NAD synthesis³¹. Thus, *SIR2* and NAD seem to be an integral part of the mechanism by which caloric restriction extends life span in yeast (Fig. 2). The activation of Sir2 by caloric restriction may occur because the slowing in metabolism frees up more of the essential cofactor, NAD³³. Consistent with this view, silencing by Sir2 was elevated under restricted conditions, resulting in a reduction in recombination in the rDNA and a stabilization of the genome³¹. Thus, Sir2 resides at the interface of metabolic rate and life span, at least in the yeast model system. It will be interesting to determine whether Sir2 proteins in higher organisms likewise mediate the benefit of caloric restriction and, more generally, couple life decisions to nutrient availability.

Another study indicates that a pathway of communication between mitochondria and the nucleus also governs yeast ageing. In several strains, the deletion of mitochondrial DNA (ρ^- mutants) was reported to increase the life span of mother cells³⁴. This increase depended on the gene *RTG2*, which has been shown to signal mitochondrial dysfunction to the nucleus in a mechanism called retrograde regulation³⁵. The retrograde pathway results in an altered pattern of nuclear gene expression in an apparent adjustment to the mitochondrial deficiency. It will be interesting to determine the importance of this pathway in normal ρ^+ yeast ageing, for example under physiological or environmental conditions in which the function of mitochondria is under stress.

Does silencing change during ageing?

Why does a strengthening of silencing by *SIR2* extend life span? One possible explanation is that normal ageing is triggered by a gradual erosion in silencing^{36,37} and that an increased expression of silencing factors forestalls these changes. In yeast, Sir2 extends life span, at least in part by silencing in the rDNA; in higher organisms, it is not clear what regions of the genome Sir2 might silence. It is possible that, as in yeast, mammalian Sir2 proteins stabilize the genome in somatic cells. In a related way, Sir2 proteins may set up or reinforce silenced domains of the genome to sculpt the pattern of gene expression in differentiated cells. In this latter scenario, a global loss of silencing may result in a loss of a robust differentiated cell phenotype, or perhaps cell death, thus contributing to the organismal changes of ageing. An alternative view is that Sir2 proteins serve a primarily regulatory function in response to nutrient deprivation by repressing expression of a few key genes that promote ageing. By setting the expression level of these genes, much like any transcription factor regulating its target genes, Sir2 would set the life span of the organism. The longevity conferred by increasing the dosage of *SIR2*, by this model, is simply due to hyper-repression of genes that cause ageing.

Hormonal control of ageing in *C. elegans*

The ageing process is also being analysed genetically in multicellular organisms, including *C. elegans*, *Drosophila* and mice. In all of these animals, single-gene mutations have been found that extend life span significantly. So far, the best characterized pathway is an insulin/insulin-like growth factor 1 (IGF-1)-like endocrine system that regulates the life span of *C. elegans* (Fig. 3). Mutations that lower the level of *daf-2*, which encodes an insulin/IGF-1 receptor homologue³⁸, cause the animal to remain active and youthful much longer than normal and to live more than twice as long^{39,40}. Elegant molecular studies have shown that DAF-2 activates a conserved phosphatidylinositol-3-OH kinase (PI(3)K)/3-phosphoinositide-dependent kinase-1 (PDK1)/Akt signal transduction pathway^{41–47} (Fig. 3a). Reduction-of-function mutations in components of this pathway, including mutations in *age-1*, which encodes a PI(3)K⁴¹, and *pdk-1*, which encodes a PDK1 homologue⁴³, extend life span^{43,48}.

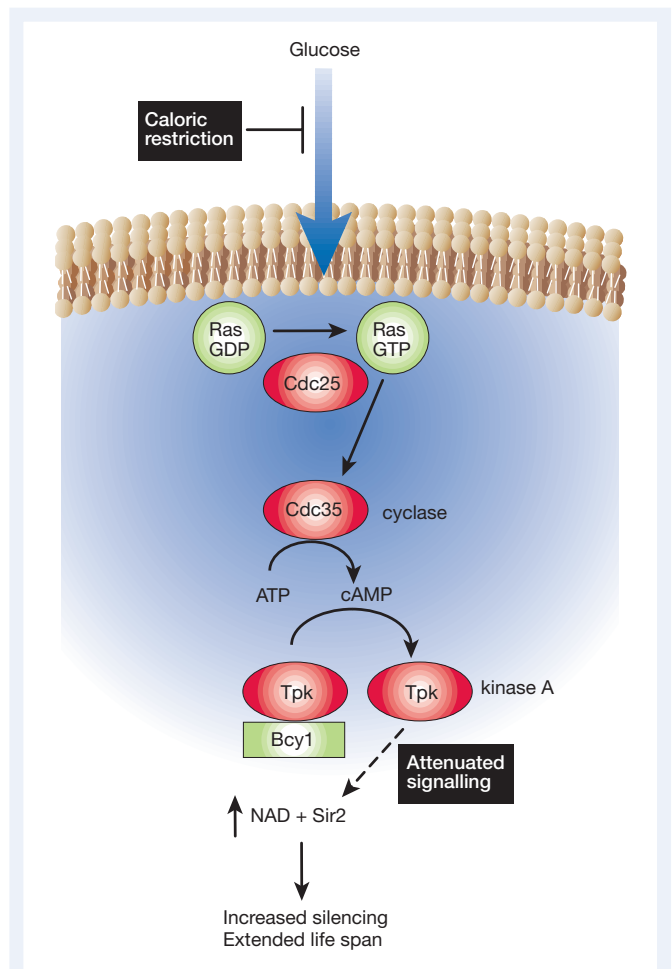


Figure 2 Caloric restriction in yeast. The carbon source glucose stimulates a signal transduction pathway including a Ras GTP-binding protein, a GTP/GDP exchange factor, an adenylate cyclase, and cAMP-dependent protein kinase A (PKA). Without cyclic AMP, PKA is in a complex with the inhibitor protein Bcy1. Signalling is attenuated either by lowering the glucose levels in the media (caloric restriction) or by mutating genes that encode components of the pathway (for example, *Cdc25*, *Cdc35* or tyrosine protein kinase (*Trk*), shown in red). This reduction in signalling leads to an increase in silencing by Sir2 and its NAD cofactor and an extended life span.

Conversely, mutations in *daf-18*, a homologue of the PTEN phosphatase^{44–47}, a negative regulator of signalling, suppress the life-span extensions of *daf-2* and *age-1* mutants^{40,49}. Recently, a candidate for a DAF-2 ligand has been identified, the insulin/IGF-1-like homologue *Ceinsulin-1*. *Ceinsulin-1* double-stranded RNA extends life span in RNA-interference experiments⁵⁰.

These long-lived mutants in the insulin/IGF-1 pathway are intriguing because many seem to have a high 'quality of life'⁵¹. Their longevity requires the activity of DAF-16 (ref. 39), a forkhead/winged-helix family transcription factor^{52,53}. Thus this signalling pathway seems likely to shorten life span by downregulating DAF-16. The discovery of DAF-16 has stimulated researchers to ask whether the vertebrate homologues of DAF-16 — AFX, FKHR and FKHL1 — function in insulin and IGF-1 signalling pathways. These homologues are indeed responsive to insulin and/or IGF-1 signalling^{54–58}, which can prevent their entry into the nucleus⁵⁵. DAF-16 contains conserved Akt-phosphorylation sites⁴², indicating that a similar mechanism might regulate DAF-16 activity in *C. elegans*.

The discovery of this pathway indicates that ageing is controlled hormonally in *C. elegans*. In some ways this is not surprising, as many other age-specific processes, such as puberty and menopause in humans, are also controlled hormonally. In principle, hormonal

regulation provides a simple way for an animal to coordinate the ageing of different tissues. In addition, it may provide a way for life span to be changed rapidly during evolution. All metazoans, with their diverse life spans, evolved from a common precursor that probably did not live very long. Thus mutations that extend life span must have been important in evolution. The evolution of body pattern seems to be driven largely by changes in regulatory genes; the same could well be true of ageing. Having a system that regulates ageing allows changes in one or a few regulatory genes to create large differences in life span. During evolution, variants with unusual life spans might be able to enter new ecological niches that favour a different rate of ageing.

In addition to regulating life span, the *C. elegans* insulin/IGF-1 signalling pathway also regulates entry into an alternative developmental state called the dauer⁵⁹. The dauer is a growth-arrested, stress-resistant alternative larval stage that is induced by food limitation and crowding. Thus dauer formation is analogous to spore formation, or its milder vertebrate equivalent, hibernation. Because it allows animals to postpone reproduction until conditions improve, dauer formation has the same consequence for *C. elegans* that the response to caloric restriction has for vertebrates. The dauer state can be considered to be a puberty checkpoint, as it arrests growth just prior to reproductive maturation. Only young larvae can become dauers; once the animals have entered 'puberty', they no longer have this option.

The mutations in the insulin/IGF-1 pathway that affect adult life span are weak mutations^{51,60}. Strong mutations in the same genes cause young larvae to arrest as dauers. This indicates that low levels of endocrine signalling are sufficient to bypass the dauer checkpoint, whereas higher levels are required for normal life span. Dauers are thought to derive energy from stored fat, which they accumulate before entry into dauer. In order for animals to remain in the dauer state, *daf-2* levels must remain low⁵⁹. Thus this metabolic shift is analogous to the metabolic shift that occurs when insulin levels fall in humans, which also triggers the breakdown and metabolism of stored fat reserves.

Why do the insulin/IGF-1-pathway mutants live so long? In addition to regulating life span, the DAF-2 receptor also regulates adult fertility and movement⁵¹. Some *daf-2* mutants (called Class 2 alleles⁵¹) adopt a dauer-like, quiescent posture, have low fertility and lay eggs very late in life, even near death^{51,60}. These fertility and movement defects can be excluded as a cause of longevity, because another class of long-lived *daf-2* mutants (class 1 alleles) move normally and have normal brood sizes and reproductive schedules⁵¹. Another possibility is that a decrease in the overall rate of metabolism lengthens life span. This is an important issue, and the findings so far are controversial. In one study, the metabolic potential of insulin/IGF-1-pathway mutants was shown to be greater than that of wild-type animals⁶¹, arguing against this model. Likewise, in a second study⁶² the metabolic rate of a long-lived *age-1* (PI(3)K) mutant was not significantly different from wild type; however, a *daf-2* allele that had the same life span as the *age-1* allele had a very low rate of metabolism. Why was this? The *daf-2* mutant examined is known to behave as a healthy class 1 allele at 20 °C but as a quiescent class 2 allele at higher temperature^{40,51,63}. In this study⁶², this mutant grew to adulthood much more slowly than others have observed at 20 °C (refs 39, 40), suggesting that it was behaving as a class 2 allele, and thus expressing physiological defects unrelated to life-span extension. It will be interesting to determine the metabolic rates of true class 1 *daf-2* alleles.

daf-2⁻ adults have dark intestinal pigmentation⁶⁴, which is correlated with a metabolic shift to fat production³⁸. However, intestinal pigmentation can be uncoupled from life span in genetic mosaic animals. *daf-2* functions non cell-autonomously to regulate both life span and intestinal metabolism⁶⁵. Some animals with *daf-2*⁻ intestines had normal intestinal pigmentation. Conversely, some animals in which part of the ectoderm was *daf-2*⁻ (but the intestine was *daf-2*⁺) had dark intestines. Many of these mosaics were long

lived. Significantly, in some classes of ectodermal mosaics, many or even all of the long-lived animals had normal intestinal pigmentation. Thus this metabolic shift is not required for life-span extension. Mutants in the insulin/IGF-1 pathway are known to be resistant to many forms of stress, including heat shock, ultraviolet (UV) light, hydrogen peroxide (H₂O₂) and paraquat⁶⁶⁻⁶⁸. It is possible that their increased ability to withstand oxidative stress is responsible for their longevity. The role of oxidative stress in ageing is discussed in more detail below.

A downstream signal

The insulin/IGF-1 pathway functions to regulate a downstream signal or hormone, because, in genetic mosaic analysis, removing *daf-2* activity from individual lineages can cause the whole animal to become a long-lived adult or to enter the dauer state⁶⁵. *daf-2* acts in lineages that produce ectodermal cells (neurons, skin), as well as lineages that produce predominantly endodermal and mesodermal cells to control both life span and dauer formation. *daf-2* also acts non cell-autonomously in the ectoderm (and also in internal tissues) to control intestinal metabolism and reproduction⁶⁵. For life span as well as dauer formation, ectodermal lineages appear to produce higher levels of downstream signal than do endodermal and mesodermal lineages. Surprisingly, loss of *daf-2* activity in tiny lineages that consist only of neurons can, at a low frequency, cause the entire animal to become a dauer. Likewise, a mosaic lacking *daf-2* only in three neurons (the ABalppap lineage) lived much longer than wild type⁶⁵. This indicates that *daf-2* functions in the nervous system to regulate dauer formation and life span. The fact that the insulin/IGF-1 pathway acts non cell-autonomously suggests that the targets of the DAF-16 transcription factor will be genes that encode, or regulate, a secreted signal (Fig. 3a,b).

Regulation of life span by sensory neurons

C. elegans can smell and taste many soluble and volatile compounds. Surprisingly, perturbations that decrease sensory perception, including mutations affecting components of sensory cilia or sensory signal-transduction pathways, as well as ablation of olfactory support cells, can extend mean life span up to 50% (ref. 69). These long-lived animals feed and reproduce normally (some actually have more progeny than normal), and they have normal rates of development. Genetic epistasis experiments indicate that sensory neurons influence life span, at least in part, by regulating the insulin/IGF-1 signalling pathway. A simple model is that an environmental signal, possibly a component of food or a pheromone, triggers the sensory neurons to secrete an insulin/IGF-1-like hormone that binds to the DAF-2 receptor and accelerates the ageing process (Fig. 3b). When sensory perception is prevented (or, in nature, when the environmental signal is absent), the hormone is not secreted, and the level of DAF-2 activity is decreased. This extends youthfulness and life span. This model is consistent with the finding that genes predicted to mediate secretion of insulin-like hormones act within the nervous system to regulate life span⁷⁰.

Regulation of life span by reproductive signals

The reproductive system, too, regulates ageing in *C. elegans*⁷¹. Killing the germ-line precursors causes the animals to live about 60% longer than normal. One might imagine that germ-line ablation extends life span simply by preventing reproduction. However, this seems unlikely because removing the entire reproductive system (the germ cells as well as surrounding somatic gonad) has no effect on life span³⁹. A nuclear hormone receptor, DAF-12 (ref. 72), is required for the longevity of germ line-ablated animals, indicating that a steroid hormone may control the change in life span. This germ-line signalling system requires *daf-16*, but it seems to act independently of the DAF-2 receptor: if the germ cells are killed in *daf-2*⁻ mutants, then the animals remain active and healthy for a long time, and live four times as long as normal. Genetic studies indicate that the

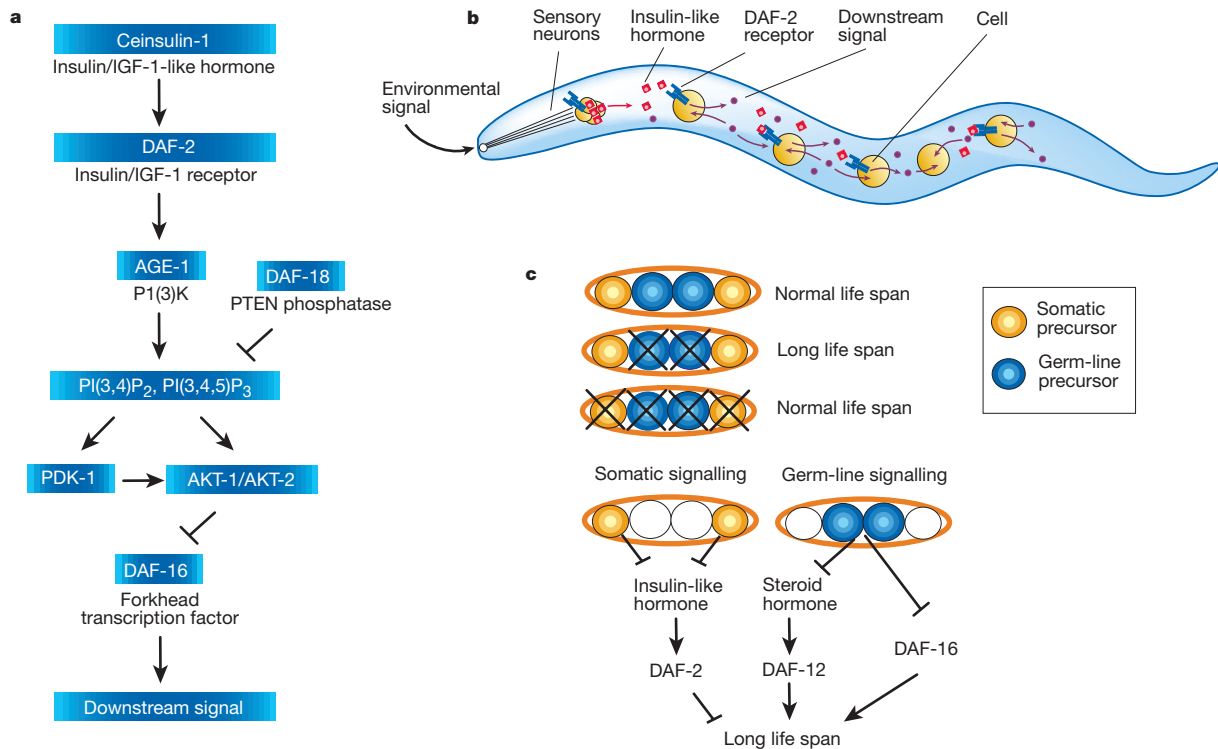


Figure 3 Regulation of *C. elegans* ageing by an elaborate endocrine system. **a**, The insulin/IGF-1-like signalling pathway that regulates ageing in *C. elegans* (see text). **b**, **c** An elaborate endocrine system involving at least four hormones seems to regulate the life span of *C. elegans*. The life span of this animal is surprisingly plastic, and is regulated by two tissues that are particularly important for its success in nature: its sensory system, which monitors environmental conditions, and its reproductive system. Both systems affect lifespan by influencing components of the DAF-2 insulin/IGF-1 pathway. **b**, A model for regulation of life span by sensory perception in *C. elegans* (see text for details). In this model, an environmental signal such as food or a pheromone causes sensory neurons to release an insulin/IGF-1-like DAF-2 ligand. Alternatively, sensory neurons could regulate the release of a DAF-2 ligand by other cells. In response to DAF-2 activity, cells either produce (shown here) or repress production of a downstream signal or hormone, which, in turn, acts through an unidentified pathway to influence aging in individual cells. The DAF-2 pathway functions in at least some neurons and probably other cells as well to regulate life span. It is not known whether DAF-2 functions in the cells that produce the insulin-like hormone. **c**, A model for regulation of life span by reproductive cells in *C. elegans* (see text and ref. 71). DAF-2, but not DAF-16, seems to function in somatic gonad signalling. Both DAF-12 and DAF-16 are required for germ-line ablation to extend life span. It is not clear whether these two proteins function in the same pathway, or in parallel pathways (as shown here). It is also not clear which non-reproductive cells function in this signalling system, although sensory neurons seem to be required for somatic gonad signalling⁷¹. Finally, sensory neurons, the insulin/IGF-1 pathway, germ cells or somatic gonad cells could activate or repress hormone production. For example, somatic gonad cells could produce an insulin/IGF-1-like hormone that acts as an antagonist.

somatic gonad produces a different signal that counterbalances the germ-line signal and extends life span, and suggest that this signal is, or regulates, a second insulin/IGF-1-like DAF-2 ligand (Fig. 3c). Surprisingly, somatic gonad signalling seems to be *daf-16* independent. Thus in the reproductive signalling system, *daf-2* can function independently of *daf-16* and vice versa.

Reproductive signalling could potentially coordinate an animal's schedule of reproduction with its rate of ageing. For example, if something were to delay the maturation of the germ line, then the animal might age more slowly. Because of this, it might remain youthful enough to bear progeny when the germ line reached maturity. Such a system could have been important in evolution, as animals with different life spans often have different reproductive schedules. Recent experiments in *Drosophila* complement these findings⁷³. Selective breeding experiments have given rise to strains of flies that produce progeny relatively late in life and are long lived, and strains that produce progeny at an earlier stage and are short lived. When the germ cells are killed, the life span of the short-lived strain is extended, and the difference between the life spans of the two strains is abolished. Thus the germ line can regulate ageing in *Drosophila*. The correlation between the time of reproduction and ageing in these two strains is intriguing. Perhaps the long-lived flies are long lived because germ-line development is delayed, causing them to produce a lower level of the germ-line 'ageing' signal.

Oxidative damage and ageing

Many findings in *C. elegans*, *Drosophila* and mice are consistent with the hypothesis that oxidative damage accelerates ageing, and that increased resistance to oxidative damage can extend life span (see review in this issue by Finkel and Holbrook, pages 239–247). So far, all of the long-lived *C. elegans* mutants tested have been found to be resistant to oxidative stress, as has a long-lived *Drosophila* mutant⁷⁴ (described below) and long-lived lines of flies produced by selective breeding^{75,76}. Mice and rats whose life spans have been extended by caloric restriction are also stress resistant⁷⁷, as are long-lived mouse mutants⁷⁸ (see below).

Conversely, mutations that increase oxidative damage can shorten life span. In *C. elegans*, mutations in *ctl-1*, a cytosolic catalase, shorten life span and prevent the life-span extension of *daf-2* mutants⁷⁹. *ctl-1* mutants seem to age more rapidly than normal, because they accumulate age-associated lipofuscin granules precociously. *ctl-1* expression also increases in *daf-2* mutants⁷⁹, as does expression of *sod-3*, which encodes a Mn-superoxide dismutase (SOD)^{61,80}. In addition, a mutation in *mev-1*, which encodes a subunit of the mitochondrial enzyme succinate dehydrogenase, accelerates the ageing process (by lipofuscin granule accumulation) and shortens life span⁸¹. This mutant exhibits increased sensitivity to oxygen. Recently, a drug known to act as an antioxidant was shown to extend the life span of *C. elegans* by up to 50% (ref. 82). In *Drosophila*,

overexpression of Cu/Zn-SOD, either ubiquitously or specifically in motor neurons, extends life span by 40–50% without affecting metabolic rate or fertility^{83,84}.

Although oxidative damage is important in ageing, it seems unlikely to be the only cause. For one thing, the magnitude of life-span extension seen in certain *C. elegans* mutants, up to fourfold greater than in wild-type animals^{40,71,85}, is much greater than that produced by treatments that counteract oxidative damage. In addition, it is possible to uncouple life-span extension from stress resistance in calorically restricted yeast³¹. Other forms of damage, as well as cellular mechanisms that repair this damage, probably influence ageing as well.

Regulation of life span by a metabolic sensor

Another intriguing system that influences the life span of *C. elegans* is the *clk* gene system (reviewed in ref. 85). Mutations in these genes slow the rates of many processes, such as cell division, the rate of 'pumping' (feeding) and defecation. In addition, they lengthen adult life span by 15–30%. Genetic experiments indicate that the *daf* and *clk* mutants probably act in different pathways, because *daf-2 clk-1* double mutants live much longer than either single mutant. In addition, the longevity of *clk-1* mutants seems to be partly independent of *daf-16*. *clk-1* is homologous to the yeast *coq7* gene, which encodes a mitochondrial protein and is involved in the synthesis of ubiquinone. The *C. elegans* gene complements the yeast mutation, indicating that it has a similar biochemical activity. Although this suggests that *clk* mutants slow the 'rate of living' because they have a low metabolic rate, *clk-1*-null mutants have nearly normal levels of mitochondrial respiration and do not seem to have reduced levels of ubiquinone.

Why are these mutants long lived? Because they live slowly, they may be long lived because their rate of metabolism is reduced, which, in turn, would be expected to reduce the rate of oxidative damage. In addition, because they pump slowly, they may be calorically restricted. Because *clk-1* mutants have normal rates of mitochondrial respiration, Hekimi and co-workers have argued that *clk-1* has a regulatory role in determining the 'rate of living'⁸⁵. Specifically, they propose that *clk-1*'s role is to report to the nucleus on the metabolic state of the mitochondria, causing nuclear genes to set the correct rates of behaviour, development and ageing. In the mutant, the nucleus does not receive this information from the mitochondria, and thus it sets the 'rate of living' too slowly. This is an interesting hypothesis, because it argues that the rate of mitochondrial respiration does not impact the 'rate of living' directly, but instead serves as a regulatory switch. This model is consistent with the observation that a small fraction of *clk-1* mutants undergo embryogenesis more rapidly than normal, suggesting deregulation rather than simple inhibition of metabolism.

Another interesting feature of this model is that it helps to explain the peculiar response of *clk-1* mutants to changes in temperature. Wild-type worms live, and age, more slowly at lower temperature than at high temperature. When two-cell embryos are transferred from high or low temperature to a moderate temperature, they promptly adjust their rate of development to the new temperature. *clk-1* mutants are unable to do this; instead, they undergo embryogenesis according to their temperature of origin. This indicates that the change in the rate of development that occurs with temperature is not simply a passive consequence of a change in thermal motion and biochemical reaction rates. Instead, this ability to conform to temperature must be under active regulation.

Life-span pathways in *C. elegans*

In *C. elegans*, at least four processes can influence life span: caloric restriction⁸⁶, signalling through the insulin/IGF-1 pathway, germ-line activity and the *clk* pathway. Although at this point it seems likely that the *clk*, insulin/IGF-1 and germ-line pathways are distinct, it is not clear whether any may function in the animal's response to caloric restriction. Certain *C. elegans eat* mutants (which do not eat

properly⁸⁷) are long lived⁸⁸. Genetic epistasis experiments indicate that these mutations and the *clk* mutations affect the same pathway⁸⁸. In contrast, *eat* mutations and *daf-2* mutations seem to affect different pathways⁸⁸. These findings suggest that the *clk* pathway, but not the insulin/IGF-1 pathway, participates in the response to caloric restriction. However, it is not clear that these *eat* mutants are actually calorically restricted, because they do not potentiate dauer formation in the way that actual food limitation does⁸⁷, and not all *eat* mutants are long lived. Caloric restriction in vertebrates causes insulin levels to fall. Thus it will be interesting to learn how mutants in the insulin/IGF-1 pathway respond to actual food limitation. Finally, overexpression of *tkr-1*, which encodes a tyrosine kinase, can extend life span in *C. elegans* by about 60%, without affecting development or fertility. These animals are also resistant to heat and UV light⁸⁹. This implies the existence of another signal-transduction pathway in the regulation of *C. elegans* life span.

A long-lived *Drosophila* mutant

Single-gene mutations can extend the life span of *Drosophila* as well as *C. elegans*. A partial loss-of-function mutation in the gene *methuselah* extends the average life span by about 35% (ref. 74). Thus in the wild type, this gene functions to shorten life span. Null alleles of *methuselah* die before reaching adulthood, indicating that the gene has an essential function during development. As mentioned earlier, *methuselah* mutants are resistant to a number of different stresses, including starvation, high temperature and paraquat, a free-radical generator. The *methuselah* gene encodes a putative G-protein-coupled seven-transmembrane-domain receptor. This suggests that a signal-transduction pathway regulates life span and stress resistance in this animal. G-protein-coupled receptors function in many types of signalling pathways. Because the *methuselah* protein has no close homologues with known function, it is difficult to predict in which type of pathway (for example, hormonal, sensory or glucose regulatory) it is likely to act.

Mutations that extend the life span of mice

A mutation in the gene encoding the protein p66^{shc} extends the life spans of mice by about 30% (ref. 78). p66^{shc} is an adaptor protein involved in the oxidative stress response. Normal cells in culture undergo apoptosis following treatment with agents that induce oxidative damage, such as H₂O₂, paraquat and UV light. In contrast, cells derived from mutant mice lacking p66^{shc} exhibit enhanced resistance to apoptosis following oxidative stress. The p66^{shc} protein is serine-phosphorylated following treatments that induce oxidative damage. This phosphorylation seems to be required for stress-induced apoptosis, because cells carrying p66^{shc} phosphorylation-site mutants are resistant to oxidative damage. The resistance to oxidative stress of p66^{shc} mutants can also be observed in whole-animal studies, as mice that lack p66^{shc} exhibit increased resistance to paraquat injection. It is not clear why p66^{shc} mutants are long lived. One possibility is that in the wild type, the loss of cells resulting from stress-induced apoptosis accelerates ageing. Because this apoptosis is blocked in p66^{shc} mutants, life span is extended. However, it is also possible that loss of p66^{shc} reduces the amount of damage induced by oxidative damaging agents in the first place. This would have the effect of increasing the health of individual cells, and it would also reduce apoptosis. In both models, additional cells would survive in the mutant mice, but in the latter case cellular components would undergo less oxidative damage. Finally, it is important to note that resistance to oxidative stress in these animals need not be the cause of their longevity. The correlation is intriguing, but not a demonstration of causality.

As in *C. elegans*, endocrine signalling seems to regulate ageing in mice. Mutations in genes that inhibit the development of the pituitary gland cause dwarfism and life-span extension. One such gene, defective in the Ames dwarf mice, encodes Prophet of Pit-1 (PROPI), a homeodomain protein that is expressed specifically in

the embryonic pituitary gland and is required for expression of Pit-1, which, in turn, is required for pituitary development⁹⁰. Ames dwarf mice lack several hormones, including growth hormone, prolactin and thyroid-stimulating hormone, and they live approximately 50% longer than normal⁹¹. Why the Ames dwarf mice are long lived is not clear. For example, it is not known whether their small body size is required for their longevity, and which hormone (or hormones) defective in the mutant may regulate life span. Interestingly, animals that overexpress growth hormone exhibit signs of accelerated ageing^{92,93}, and growth hormone-receptor mutants seem to have extended life spans⁹². Like calorically restricted mice, the dwarf mice have a low body temperature⁹⁴. In addition, they have elevated levels of catalase and Cu/Zn SOD, and decreased levels of reactive oxygen species^{95,96}. These features are consistent with their longevity being due to reduced metabolic rate or oxidative damage. However, a report investigating the level of these defence proteins in the hypothalamus⁹⁶ complicates this interpretation. In this tissue, both the long-lived dwarf mice as well as the short-lived mice overexpressing growth hormone have increased levels of catalase and Cu/Zn SOD.

Conservation

Many biological processes are conserved widely throughout the animal kingdom. Is this true of ageing? Caloric restriction can extend life span in a wide range of organisms; thus this seems like a good candidate for an evolutionarily conserved ageing pathway. Little is known about the genes that regulate life-span extension in response to caloric intake. In this regard, it will be particularly interesting to learn whether the yeast SIR2 histone deacetylase pathway (described above) regulates ageing in metazoans and, if so, whether it acts in the response to caloric restriction. In addition to caloric restriction, there are indications that both germ-line activity and oxidative damage influence ageing in both flies and worms, and that oxidative damage might influence ageing in vertebrates as well (see review in this issue by Finkel and Holbrook, pages 239–247). In addition, signalling from the mitochondria may regulate life span in at least two organisms, *C. elegans* and yeast. Does insulin/IGF-1 signalling regulate life span in animals other than *C. elegans*? Mutations affecting insulin/IGF-1 signalling have been discovered in flies, but their effect on life span is not known. In humans, insulin-receptor mutations cause diabetes rather than longevity; however, these mutations lead ultimately to a severe loss of insulin production and so may not mimic the weak signalling mutations of *C. elegans*. Further complicating the situation, vertebrates have at least three different insulin-receptor family members, and it is becoming clear that different branches of the insulin/IGF-1 signalling network control different aspects of physiology and metabolism. Possibly one branch of this network controls ageing. This seems particularly plausible given the effects of endocrine perturbation on ageing in dwarf mice, especially since the dwarf mice have reduced levels of insulin⁹⁵. If ageing does prove to be regulated hormonally in humans, then it may be possible to extend youthfulness and increase the quality of old age with drugs or hormones that perturb the endocrine system.

The long-lived *Drosophila methuselah* mutant, which is defective in a putative G-protein-coupled receptor, indicates that cell–cell communication or endocrine signalling may influence the life span of flies, but it is not clear whether *methuselah* homologues regulate life span in other animals. A similar situation exists with the *C. elegans tkr-1* gene, which encodes a tyrosine kinase receptor that appears to extend life span⁸⁹. Ultimately it seems likely that each known life-span gene will be found to act in one of several pathways that regulate ageing, and that at least some of these pathways will be highly conserved. These are still early days; the screens for life-span mutants are not saturated in any organism.

Diseases of ageing

Ageing in humans is manifest not only by the stereotypical changes in phenotype but also by a large increase in the onset of many diseases.

Can these diseases be thought of as an integral part of the ageing process; that is, would some anti-ageing intervention slow their onset? A useful example to consider may be cancer. Clearly, like ageing, the progression of many cancers is a time-dependent process. The role of somatic mutations in the pathways of cancer development has been amply documented (see review in this issue by DePinho, pages 248–254). Therefore, the development of at least certain cancers would seem to be mechanistically separate from ageing. However, a link between ageing and cancer is suggested by the delay of cancers in calorically restricted mice, and by the striking correlation between physiological age and the likelihood of cancer in animals with very different life spans. This suggests that any treatment that slows the ageing process, whether it acts directly on hormone signalling, gene silencing or oxidative stress, may have the potential to delay cancer and possibly other diseases of ageing.

Summary

The field of ageing research has been completely transformed in the past decade. It is now widely accepted that the ageing process, like most biological processes, is subject to regulation and can be studied using classical genetics. When single genes are changed, animals that should be old stay young. In humans, these mutants would be analogous to a ninety year old who looks and feels forty-five. On this basis we begin to think of ageing as a disease that can be cured, or at least postponed. This paradigm shift is due largely to the analysis of single-gene mutations that influence ageing in model organisms. The field of ageing is beginning to explode, because so many are so excited about the prospect of searching for — and finding — the causes of ageing, and maybe even the fountain of youth itself.

Note added in proof. Recently, Wolkow *et al.*⁹⁷ have shown that expression of the *C. elegans daf-2* receptor or *age-1* PI(3)K only in neurons can confer normal life span, and that expression of *daf-2* only in endoderm has a significant, but lesser, effect. These findings are in accord with previous mosaic analysis⁶⁵. An important caveat is that the levels of DAF-2 and AGE-1 produced in these transgenic animals may differ from endogenous levels, possibly altering the level of downstream signal. Wolkow *et al.* also demonstrate that expressing *daf-2* or *age-1* only in neurons can be sufficient for normal fat metabolism in the intestine. This is consistent with earlier findings that *daf-2* activity in the ectoderm can be necessary and sufficient for normal intestinal pigmentation⁶⁵, although it should be noted that genetic mosaic animals with a nervous system that is almost completely wild type but internal tissues that are *daf-2*[−] often have a *Daf-2*[−] intestinal phenotype. Finally, the new study reported that the altered intestinal metabolism of insulin/IGF-1 pathway mutants is not required for longevity, as had been shown previously by analysing intestinal pigmentation and life span in genetic mosaics⁶⁵ (see text). □

- Weindruch, R. H., Walford, R. L., Fligiel, S. & Guthrie, D. The retardation of aging in mice by dietary restriction: longevity, cancer, immunity, and lifetime energy intake. *J. Nutr.* **116**, 641–654 (1986).
- Harman, D. The aging process. *Proc. Natl Acad. Sci. USA* **78**, 7124–7128 (1981).
- Finch, C. *Longevity, senescence, and the genome* (Univ. Chicago Press, Chicago, IL, 1990).
- Mortimer, R. K. & Johnston, J. R. Life span of individual yeast cells. *Nature* **183**, 1751–1752 (1959).
- Egilmez, N. K. & Jazwinski, S. M. Evidence for the involvement of a cytoplasmic factor in the aging of the yeast *Saccharomyces cerevisiae*. *J. Bacteriol.* **171**, 37–42 (1989).
- Jazwinski, S. M. The genetics of aging in the yeast *Saccharomyces cerevisiae*. *Genetics* **91**, 35–53 (1993).
- Smeal, T., Claus, J., Kennedy, B., Cole, F. & Guarente, L. Loss of transcriptional silencing causes sterility in old mother cells of *S. cerevisiae*. *Cell* **84**, 633–642 (1996).
- Rine, J. & Herskowitz, I. Four genes responsible for a position effect on expression from HML and HMR in *Saccharomyces cerevisiae*. *Genetics* **116**, 9–22 (1987).
- Gottschling, D. E., Aparicio, O. M., Billington, B. L. & Zakian, V. A. Position effect at *S. cerevisiae* telomeres: reversible repression of Pol II transcription. *Cell* **63**, 751–762 (1990).
- Kennedy, B. K., Austriaco, N. R. Jr, Zhang, J. & Guarente, L. Mutation in the silencing gene SIR4 can delay aging in *S. cerevisiae*. *Cell* **80**, 485–486 (1995).
- Smith, J. S. & Boeke, J. D. An unusual form of transcriptional silencing in yeast ribosomal DNA. *Genes Dev.* **11**, 241–254 (1997).
- Bryk, M. *et al.* Transcriptional silencing of Ty1 elements in the RDN1 locus of yeast. *Genes Dev.* **11**, 255–269 (1997).
- Kaeberlein, M., McVey, M. & Guarente, L. The SIR2/3/4 complex and SIR2 alone promote longevity in *Saccharomyces cerevisiae* by two different mechanisms. *Genes Dev.* **13**, 2570–2580 (1999).

14. Kim, S., Benguria, A., Lai, C.-Y. & Jazwinski, M. Modulation of life span by histone deacetylase genes in *S. cerevisiae*. *Mol. Biol. Cell* **10**, 3125–3136 (1999).
15. Smith, J. S., Caputo, E. & Boeke, J. A genetic screen for ribosomal DNA silencing defects identifies multiple DNA replication and chromatin remodeling factors. *Mol. Cell. Biol.* **19**, 3184–3197 (1999).
16. Gottlieb, S. & Esposito, R. E. A new role for a yeast transcriptional silencer gene, SIR2, in regulation of recombination in ribosomal DNA. *Cell* **56**, 771–776 (1989).
17. Sinclair, D. & Guarente, L. Extrachromosomal rDNA circles—a cause of aging in yeast. *Cell* **91**, 1033–1042 (1997).
18. Defossez, P. A. *et al.* Elimination of replication block protein Fob1 extends the life span of yeast mother cells. *Mol. Cell* **3**, 447–455 (1999).
19. Heo, S. T., K., Ohsugi, I., Shimamoto, A., Furiuchi, Y. & Ikeda, I. Blooms syndrome gene suppresses premature aging causes by Sgs1 deficiency in yeast. *Genes Cells* **4**, 619–624 (1999).
20. Ashrafi, K., Sinclair, D., Gordon, J. & Guarente, L. Passage through stationary phase advances replicative aging in *S. cerevisiae*. *Proc. Natl Acad. Sci. USA* **96**, 9100–9105 (1999).
21. Sinclair, D., Mills, K. & Guarente, L. Accelerated aging and nucleolar fragmentation in yeast sgs1 mutants. *Science* **277**, 1313–1316 (1997).
22. Celenza, J. L. A yeast gene that is essential for release from glucose repression encodes a protein kinase. *Science* **233**, 1175–1180 (1986).
23. Ashrafi, K., Lin, S., Manchester, J. & Gordon, J. Sip2p and its partner Snf1p kinase affect aging in *S. cerevisiae*. *Genes Dev.* **14**, 1872–1885 (2000).
24. Brachmann, C. B. *et al.* The SIR2 gene family, conserved from bacteria to humans, functions in silencing, cell cycle progression, and chromosome stability. *Genes Dev.* **9**, 2888–2902 (1995).
25. Frye, R. A. Characterization of five human cDNAs with homology to yeast SIR2 gene: Sir2-like proteins (Sirtuins) metabolize NAD and may have protein ADP-ribosyltransferase activity. *Biochem. Biophys. Res. Commun.* **260**, 273–279 (1999).
26. Braustein, M., Rose, A. B., Holmes, S. G., Allis, C. D. & Broach, J. R. Transcriptional silencing in yeast is associated with reduced nucleosome acetylation. *Genes Dev.* **7**, 592–604 (1993).
27. Tanny, J. C., Dowd, G. J., Huang, J., Hilz, H. & Moazed, D. An enzymatic activity in the yeast SIR2 protein that is essential for gene silencing. *Cell* **99**, 735–745 (1999).
28. Imai, S. I., Armstrong, C. M., Kaerberlein, M. & Guarente, L. Transcriptional silencing and longevity protein SIR2 is an NAD-dependent histone deacetylase. *Nature* **403**, 795–799 (2000).
29. Smith, J. S. *et al.* A phylogenetically conserved NAD⁺-dependent protein deacetylase activity in the Sir2 protein family. *Proc. Natl Acad. Sci. USA* **97**, 6658–6663 (2000).
30. Landry, J. *et al.* The silencing protein SIR2 and its homologs are NAD-dependent protein deacetylases. *Proc. Natl Acad. Sci. USA* **97**, 5807–5811 (2000).
31. Lin, S., Defossez, P. & Guarente, L. Life span extension by calorie restriction in *S. cerevisiae* requires NAD and SIR2. *Science* **289**, 2126–2128 (2000).
32. Chen, J. B., Sun, J. & Jazwinski, S. M. Prolongation of the yeast life span by the v-Ha-Ras oncogene. *Mol. Microbiol.* **4**, 2081–2086 (1990).
33. Guarente, L. Sir2 links chromatin silencing, metabolism, and aging. *Genes Dev.* **14**, 1021–1026 (2000).
34. Kirchman, P. A., Sangkyu, K., Lai, C. Y. & Jazwinski, S. M. Interorganellar signaling is a determinant of longevity in *Saccharomyces cerevisiae*. *Genetics* **152**, 179–190 (1999).
35. Parikh, V. S., Morgan, M., Scott, R., Clements, S. & Butow, R. A. The mitochondrial genotype can influence nuclear gene expression in yeast. *Science* **235**, 576–580 (1987).
36. Guarente, L. Do changes in chromosomes cause aging? *Cell* **86**, 9–12 (1996).
37. Wareham, K. A., Lyon, M. F., Glenister, P. H. & Williams, E. D. Age related reactivation of an X-linked gene. *Nature* **327**, 725–727 (1987).
38. Kimura, K. D., Tissenbaum, H. A., Liu, Y. & Ruvkun, G. *daf-2*, an insulin receptor-like gene that regulates longevity and diapause in *Caenorhabditis elegans*. *Science* **277**, 942–946 (1997).
39. Kenyon, C., Chang, J., Gensch, E., Rudner, A. & Tabtiang, R. A *C. elegans* mutant that lives twice as long as wild type. *Nature* **366**, 461–464 (1993).
40. Larsen, P. L., Albert, P. S. & Riddle, D. L. Genes that regulate both development and longevity in *Caenorhabditis elegans*. *Genetics* **139**, 1567–1583 (1995).
41. Morris, J. Z., Tissenbaum, H. A. & Ruvkun, G. A phosphatidylinositol-3-OH kinase family member regulating longevity and diapause in *Caenorhabditis elegans*. *Nature* **382**, 536–539 (1996).
42. Paradis, S. & Ruvkun, G. *Caenorhabditis elegans* Akt/PKB transduces insulin receptor-like signals from AGE-1 PI3 kinase to the DAF-16 transcription factor. *Genes Dev.* **12**, 2488–2498 (1998).
43. Paradis, S., Ailion, M., Toker, A., Thomas, J. H. & Ruvkun, G. A PDK1 homolog is necessary and sufficient to transduce AGE-1 PI3 kinase signals that regulate diapause in *Caenorhabditis elegans*. *Genes Dev.* **13**, 1438–1452 (1999).
44. Ogg, S. & Ruvkun, G. The *C. elegans* PTEN homolog, DAF-18, acts in the insulin receptor-like metabolic signaling pathway. *Mol. Cell* **2**, 887–893 (1998).
45. Gil, E. B., Link, E. M., Liu, L. X., Johnson, C. D. & Lees J. A. Regulation of the insulin-like developmental pathway of *Caenorhabditis elegans* by a homolog of the PTEN tumor suppressor gene. *Proc. Natl Acad. Sci. USA* **96**, 2925–2930 (1999).
46. Rouault, J. P. *et al.* Regulation of dauer larva development in *Caenorhabditis elegans* by *daf-18*, a homologue of the tumour suppressor PTEN. *Curr. Biol.* **9**, 329–332 (1999).
47. Mihaylova, V. T., Borland, C. Z., Manjarrez, L., Stern, M. J. & Sun, H. The PTEN tumor suppressor homolog in *Caenorhabditis elegans* regulates longevity and dauer formation in an insulin receptor-like signaling pathway. *Proc. Natl Acad. Sci. USA* **96**, 7427–7432 (1999).
48. Friedman, D. B. & Johnson, T. E. Three mutants that extend both mean and maximum life span of the nematode, *Caenorhabditis elegans*, define the *age-1* gene. *J. Gerontol.* **43**, 102–109 (1988).
49. Dorman, J. B., Albinder, B., Shroyer, T. & Kenyon, C. The *age-1* and *daf-2* genes function in a common pathway to control the lifespan of *Caenorhabditis elegans*. *Genetics* **141**, 1399–1406 (1995).
50. Kawano, T. *et al.* Molecular cloning and characterization of a new insulin/IGF-like peptide of the nematode *Caenorhabditis elegans*. *Biochem. Biophys. Res. Commun.* **273**, 431–436 (2000).
51. Gems, D. *et al.* Two pleiotropic classes of *daf-2* mutation affect larval arrest, adult behavior, reproduction and longevity in *Caenorhabditis elegans*. *Genetics* **150**, 129–155 (1998).
52. Lin, K., Dorman, J. B., Rodan, A. & Kenyon, C. *daf-16*: An HNF-3/forkhead family member that can function to double the life-span of *Caenorhabditis elegans*. *Science* **278**, 1319–1322 (1997).
53. Ogg, S. *et al.* The Fork head transcription factor DAF-16 transduces insulin-like metabolic and longevity signals in *C. elegans*. *Nature* **389**, 994–999 (1997).
54. Nakae, J., Park, B. & Accili, D. Insulin stimulates phosphorylation of the Forkhead transcription factor FKHR on serine 253 through a wortmannin-sensitive pathway. *J. Biol. Chem.* **274**, 15982–15985 (1999).
55. Brunet, A. *et al.* Akt promotes cell survival by phosphorylating and inhibiting a Forkhead transcription factor. *Cell* **96**, 857–868 (1999).
56. Rena, G., Guo, S., Cichy, S. C., Unterman, T. G. & Cohen, P. Phosphorylation of the transcription factor Forkhead family member FKHR by protein kinase B. *J. Biol. Chem.* **274**, 17179–17183 (1999).
57. Kops, G. J. *et al.* Direct control of the Forkhead transcription factor AFX by protein kinase B. *Nature* **398**, 630–634 (1999).
58. Biggs, W. H. III, Meisenhelder, J., Hunter, T., Cavenee, W. K. & Arden, K. C. Protein kinase B/Akt-mediated phosphorylation promotes nuclear exclusion of the winged helix transcription factor FKHR1. *Proc. Natl Acad. Sci. USA* **96**, 7421–7460 (1999).
59. Riddle, D. L. & Albert, P. S. in *C. elegans* Vol. II (eds Riddle, D. L., Blumenthal, T., Meyer, B. J. & Priess, J. R.) 739–768 (Cold Spring Harbor Laboratory Press, 1997).
60. Tissenbaum, H. A. & Ruvkun, G. An insulin-like signaling pathway affects both longevity and reproduction in *Caenorhabditis elegans*. *Genetics* **148**, 703–717 (1998).
61. Vanfleteren, J. R. & De Vreese, A. The gerontogenes *age-1* and *daf-2* determine metabolic rate potential in aging *Caenorhabditis elegans*. *FASEB J.* **9**, 1355–1361 (1995).
62. Van Voorhies, W. A. & Ward, S. Genetic and environmental conditions that increase longevity in *Caenorhabditis elegans* decrease metabolic rate. *Proc. Natl Acad. Sci. USA* **96**, 11399–11403 (1999).
63. Gottlieb, S. & Ruvkun, G. *daf-2*, *daf-16* and *daf-23*: genetically interacting genes controlling dauer formation in *Caenorhabditis elegans*. *Genetics* **137**, 107–120 (1994).
64. Vowels, J. J. & Thomas, J. H. Genetic analysis of chemosensory control of dauer formation in *C. elegans*. *Genetics* **130**, 105–123 (1992).
65. Apfeld, J. & Kenyon, C. Cell nonautonomy of *C. elegans* DAF-2 function in the regulation of diapause and life span. *Cell* **95**, 199–210 (1998).
66. Larsen, P. L. Aging and resistance to oxidative damage in *Caenorhabditis elegans*. *Proc. Natl Acad. Sci. USA* **90**, 8905–8909 (1993).
67. Lithgow, G. J., White, T. M., Melov, S. & Johnson, T. E. Thermotolerance and extended life-span conferred by single-gene mutations and induced by thermal stress. *Proc. Natl Acad. Sci. USA* **92**, 7540–7544 (1995).
68. Murakami, S. & Johnson, T. E. A genetic pathway conferring life extension and resistance to UV stress in *Caenorhabditis elegans*. *Genetics* **143**, 1207–1218 (1996).
69. Apfeld, J. & Kenyon, C. Regulation of lifespan by sensory perception in *Caenorhabditis elegans*. *Nature* **402**, 804–809 (1999).
70. Ailion, M., Inoue, T., Weaver, C. I., Holdcraft, R. W. & Thomas, J. H. Neurosecretory control of aging in *Caenorhabditis elegans*. *Proc. Natl Acad. Sci. USA* **96**, 7394–7397 (1999).
71. Hsin, H. & Kenyon, C. Signals from the reproductive system regulate the lifespan of *C. elegans*. *Nature* **399**, 362–366 (1999).
72. Antebi, A., Yeh, W., Tait, D., Hedgecock, E. M. & Riddle, D. L. *daf-12* encodes a nuclear receptor that regulates the dauer diapause and developmental age in *C. elegans*. *Genes Dev.* **14**, 1512–1527 (2000).
73. Sgro, C. M. & Partridge, L. A delayed wave of death from reproduction in *Drosophila*. *Science* **286**, 2521–2524 (1999).
74. Lin, Y. J., Seroude, L. & Benzer, S. Extended life-span and stress resistance in the *Drosophila* mutant *methuselah*. *Science* **282**, 943–946 (1998).
75. Arking, R., Buck, S., Berrios, A., Dwyer, S. & Baker, G. T. Elevated paraquat resistance can be used as a bioassay for longevity in a genetically based long-lived strain of *Drosophila*. *Dev. Genet.* **12**, 362–370 (1991).
76. Service, P. M., Hutchinson, E. W., MacKinley, M. D. & Rose, M. R. Resistance to environmental stress in *Drosophila melanogaster* selected for postponed senescence. *Physiol. Zool.* **58**, 380–389 (1985).
77. Sohal, R. S. & Weindruch, R. Oxidative stress, caloric restriction, and aging. *Science* **273**, 59–63 (1996).
78. Migliaccio, E. *et al.* The p66^{cdc42} adaptor protein controls oxidative stress response and life span in mammals. *Nature* **402**, 309–313 (1999).
79. Taub, J. *et al.* A cytosolic catalase is needed to extend adult lifespan in *C. elegans daf-2* and *clk-1* mutants. *Nature* **399**, 162–166 (1999).
80. Honda, Y. & Honda, S. The *daf-2* gene network for longevity regulates oxidative stress resistance and Mn-superoxide dismutase gene expression in *Caenorhabditis elegans*. *FASEB J.* **13**, 1385–1393 (1999).
81. Ishii, N. *et al.* A mutation in succinate dehydrogenase cytochrome *b* causes oxidative stress and ageing in nematodes. *Nature* **394**, 694–697 (1998).
82. Melov, S. *et al.* Extension of life-span with superoxide dismutase/catalase mimetics. *Science* **289**, 1567–1569 (2000).
83. Sun, J. & Tower, J. FLP recombinase-mediated induction of Cu/Zn-superoxide dismutase transgene expression can extend the life span of adult *Drosophila melanogaster* flies. *Mol. Cell. Biol.* **19**, 216–228 (1999).
84. Parkes, T. L. *et al.* Extension of *Drosophila* lifespan by overexpression of human SOD1 in motorneurons. *Nature Genet.* **19**, 171–174 (1998).
85. Branicky, R., Benard, C. & Hekimi, S. *clk-1*, mitochondria, and physiological rates. *BioEssays* **22**, 48–56 (2000).
86. Klass, M. R. Aging in the nematode *C. elegans*: major biological and environmental factors influencing life span. *Mech. Ageing Dev.* **6**, 413–429 (1977).
87. Avery, L. The genetics of feeding in *Caenorhabditis elegans*. *Genetics* **133**, 897–917 (1993).
88. Lakowski, B. & Hekimi, S. The genetics of caloric restriction in *Caenorhabditis elegans*. *Proc. Natl Acad. Sci. USA* **95**, 13091–13096 (1998).
89. Murakami, S. & Johnson, T. E. Life extension and stress resistance in *Caenorhabditis elegans* modulated by the *tkr-1* gene. *Curr. Biol.* **9**, 791–795 (1998).
90. Wu, W. *et al.* Mutations in PRO1 cause familial combined pituitary hormone deficiency. *Nature Genet.* **18**, 147–149 (1998).
91. Brown-Borg, H. M., Borg, K. E., Meliska, C. J. & Bartke, A. Dwarf mice and the aging process. *Nature* **384**, 33 (1996).
92. Bartke, A. *et al.* Does growth hormone prevent or accelerate aging? *Exp. Gerontol.* **33**, 675–687 (1998).
93. Kopchick, J. J. & Larone, Z. Is the Laron mouse an accurate model of Laron syndrome? *Molec. Gen.* **384**, 232–236 (1999).
94. Hunter, W. S., Croson, W. B., Bartke, A., Gentry, M. V. & Meliska, C. J. Low body temperature in long-lived Ames dwarf mice at rest and during stress. *Physiol. Behav.* **67**, 433–437 (1999).
95. Brown-Borg, H. M. & Rakoczy, S. G. Catalase expression in delayed and premature aging mouse models. *Exp. Gerontol.* **35**, 199–212 (1999).
96. Hauck, S. J. & Bartke, A. Effects of growth hormone on hypothalamic catalase and Cu/Zn superoxide dismutase. *Free Radical Biol. Med.* **28**, 970–978 (2000).
97. Wolkow, C. A., Kimura, K. D., Lee, M.-S. & Ruvkun, G. Regulation of *C. elegans* life-span by insulin like signaling in the nervous system. *Science* **290**, 147–150 (2000).

Lessons from human progeroid syndromes

George M. Martin*† & Junko Oshima*

Departments of Pathology* and Genetics†, University of Washington, Seattle, Washington 98195, USA

A number of human genes have been identified in which mutations can lead to the accelerated emergence of features of senescence. Studies of these genes, and of the functions of their protein products, may lead to a clearer understanding of the nature of senescence, and could provide clues for ways in which ageing might be retarded.

Ageing, or senescence, results from a waning of the force of natural selection with respect to the age of gene effects¹. The evolutionary theory of ageing predicts that a large array of mutations and polymorphisms could potentially impact upon one or more senescent phenotypes. For diseases like dementias of the Alzheimer type, only a single tissue is impacted (the brain), and we can therefore refer to these as 'unimodal progeroid syndromes'². But there is also evidence for a much smaller subset of single-gene mutations that impact, with varying fidelity, upon multiple aspects of the complex senescent phenotype, prompting us to label these as 'segmental progeroid syndromes'². They include mutations resulting in genomic instability, alterations in lipid and carbohydrate metabolism, a triplet-repeat disorder (myotonic dystrophy) and an idiopathic disorder, the Hutchinson–Gilford syndrome (HGS). It is an oversimplification to refer to these diseases as 'premature ageing syndromes'. To do so suggests that they will invariably reveal the mechanisms underlying 'usual' or 'normal' ageing. But their phenotypic features may sometimes be quite unusual, and some features may result from gross abnormalities in development. Although gene action in development may have consequences for post-maturational ageing, gerontologists are concerned primarily with differential modulation of the rates of post-maturational ageing following normal development. Most segmental progeroid syndromes do not in fact escape the force of natural selection and thus would not satisfy the evolutionary theory of gene action in ageing¹. A rationale for studying them, however, is that there may be 'leaky' alleles and polymorphisms at these loci such that the resulting phenotypes do escape the force of natural selection. Moreover, certain 'antigeroid' alleles at these loci may contribute to unusually robust retention of structure and function during ageing.

Werner syndrome and genomic instability

In the normal ageing process, malignant neoplasms are the most obvious phenotypic examples related to the loss of genomic stability during ageing. They are conspicuous because of their clonal amplifications. But somatic mutation may be more widespread among various non-neoplastic cell types in ageing mammals. Age-related increases in the frequencies of mutations, for example, have been documented in human lymphocytes³ and in renal tubular epithelial cells⁴. There is also the potential for

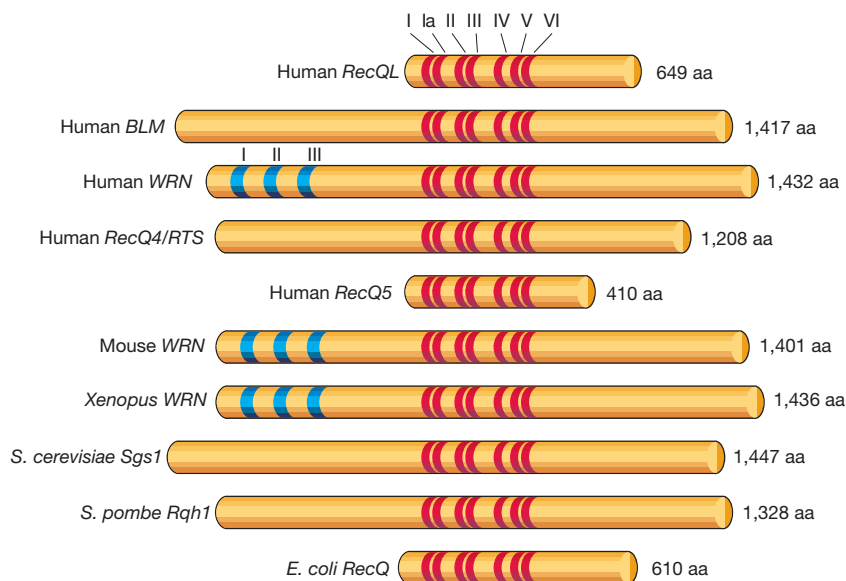
unrepaired bulky DNA adducts to accumulate and to interfere with transcription in obligate postreplicative cell types such as neurons. The striking increases in the prevalence of trisomic products of meiosis in women over the age of 35 years provides evidence that the DNA of the germ line, as well as the DNA of the soma, is vulnerable during ageing. And the paternal germ line does not escape — older fathers are at increased risk of producing mutant offspring⁵.

The genomic instability syndromes include a relatively large group of mutations and thus highlight the importance of DNA transactions (such as replication, repair, transcription and recombination) for research on ageing and diseases of ageing (Table 1). The prototype is the Werner syndrome, caused by homozygosity for null mutations in a member of the RecQ family of helicases⁶. Cells from affected individuals have been shown to have high somatic mutation rates, particularly deletions⁷.

Parents of Werner syndrome patients typically report nothing unusual about offspring afflicted with that disorder until about the time of puberty, when they note a failure to undergo the usual adolescent growth spurt⁸. There is no 'catch-up' growth, with the consequence that short stature is one of the key features of the disorder. Premature greying and thinning of hair, atrophy of skin, regional atrophy of subcutaneous tissue, voice changes (weak, high-pitched) and diminished fertility are already apparent during the third decade. Premature testicular atrophy has been well documented in specimens from middle-aged males and there is probably an accelerated loss of primordial ovarian follicles. By the beginning of the fourth decade, most patients require surgical excision of bilateral cataracts; the eventual penetrance of that phenotype is close to 100%. There then follows a variable course, with the most constant feature being the eventual development of severe ulcerations around the Achilles' tendons and malleoli (around ankles). Other features include osteoporosis (fragile bones arising from loss of bone tissue), type 2 diabetes mellitus, a variety of benign and malignant neoplasms and several types of arteriosclerosis (loss of elasticity and thickening of the walls of the arteries), including medial calcinosis (calcium deposition in blood vessels), arteriolosclerosis of small arteries, and atherosclerosis (accumulation of lipids and fibrous elements in the large arteries). The cause of death is either myocardial infarction (the consequence of coronary artery atherosclerosis) or cancer, with a median age of death of around 47–48 years.

A careful examination of the phenotypic details of Werner patients reveals unusual features. First, the ratio of

Figure 1 RecQ helicases. *WRN* homologues of human, mouse, *Xenopus*, yeasts (*Saccharomyces cerevisiae* and *Schizosaccharomyces pombe*) and *Escherichia coli* are shown. Red boxes indicate helicase domains I–IV, blue boxes indicate exonuclease domains I–III.



sarcomas to carcinomas is around 1:1, compared with 1:10 in the general ageing population⁹, indicating a special vulnerability of mesenchymal cells. Meningiomas and certain rare melanomas are particularly prevalent. The only frequent epithelial neoplasm is follicular carcinoma of the thyroid, at least in Japanese subjects. Second, the distribution of the osteoporosis is unusual; the long bones of the lower limbs can be more severely affected than those of the vertebral column. Other unusual radiological features include a characteristic osteosclerosis of the distal phalanges and subcutaneous calcification of the soft tissues. The ulcerations mentioned above are also unusual and can involve the skin around the elbows as well as the ankles. The pathogenesis seems not to be related to diabetes or to vascular insufficiency, and may instead reflect an unusual response to repeated, mild local trauma. There is no convincing evidence of premature senescence in the central nervous system (CNS), although CNS lesions may be secondary to the vascular pathology. Skeletal muscle atrophy is at least in part due to disuse, but a primary involvement of that tissue cannot yet be ruled out.

Although we were not surprised to find that an abnormality in DNA metabolism could explain the susceptibility of Werner syndrome patients to neoplasia, that such an abnormality could be the basis for all of the other lesions was intriguing. Many investigators have emphasized post-translational alterations of long-lived crystalline proteins as the basis for senescent ocular cataracts. It is apparent in Werner syndrome that the cataracts result from alterations in the lens epithelial cells. Of most interest, however, is the coupling of an abnormality in a RecQ helicase to severe atherosclerosis. In analogy with the skin ulcers seen in Werner subjects, perhaps normal Werner helicase function is required for the efficient repair of the haemodynamic shear stress of arteries¹⁰. Such repair could be at the level of endothelial cell replication; a limited replicative potential of somatic cells from Werner syndrome patients is well documented¹¹. A key question is the extent to which more subtle variations in the function of Werner helicase may contribute to differing susceptibilities to atherosclerosis. First impressions are that this would seem unlikely, as heterozygous carriers do not obviously suffer from premature atherosclerosis. But the question remains cogent because of evidence, from family-based studies, that somatic cells from heterozygotes display levels of sensitivity to a genotoxic agent (4-nitroquinoline-1-oxide) that is intermediate between that of wild-type and homozygous deficient subjects¹². A number of polymorphic variants of the Werner helicase have been elucidated¹³ and a single case-control study associating a polymorphic allele with

relative resistance to myocardial infarction in a Japanese population awaits confirmation¹⁴. The same is true for a study implicating an enrichment of an allele in a centenarian population¹⁵. More definite studies must await an evaluation of haplotypes and functional studies.

The most interesting feature of the Werner helicase, in comparison with other members of the RecQ family, is the presence of an exonuclease domain in the amino-terminal portion of the molecule (Fig. 1)^{15–17}. There is evidence that it functions both as a 5'→3' and 3'→5' exonuclease. Since the cloning of the Werner helicase gene in 1996⁶, there have been reports that it functions in replication^{18–20}, repair^{12,15–17,21–23}, transcription^{24,25} and recombination^{26,27}. It is too early to determine which of these several potential DNA transactions are the key(s) to understanding the pathogenesis of Werner syndrome and, possibly, of some aspects of normal ageing. WRN could well be involved in disrupting DNA structures formed by the illegitimate interactions between DNA strands during all of the above DNA transactions¹⁷, thus maintaining overall genomic stability.

The triplet-repeat disorder myotonic dystrophy

Myotonic dystrophy, also known as dystrophia myotonica or Steinert disease, is caused by autosomal dominant mutations at two distinct loci, *DM1* (on chromosome 19) and *DM2* (on chromosome 3). The two are not readily distinguishable phenotypically, although further study of the recently discovered *DM2* disorder may eventually reveal such distinctions. Although muscle weakness and wasting is a particularly conspicuous clinical feature of the disease, cataracts, frontal balding, testicular atrophy, some degree of cognitive decline, myocardial conduction defects and cholelithiasis are part of the syndrome. These are all features sometimes found within the general ageing population. *DM1* results from an amplification of strings of CTG repeats at the 3' untranslated region of a gene with features of a new class of a multidomain protein kinase²⁸. The kinase may be important in modulating cell size and shape. It has been suggested that the CTG repeats are essential *cis*-acting elements for a splicing event; polymorphisms may therefore influence the ratios of various isoforms, the functions of which remain unknown²⁹.

There is a wide range of allelic variants characterized by differing lengths of CTG repeats. Analysis of a heterogeneous group of patients has indicated an inverse correlation of repeat lengths with ages of onset³⁰. But a more recent analysis indicates that this correlation applies for patients with relatively small expansions, but not for the subset having large expansions³¹. Meiotic instability of loci with expanded repeats may lead to the generation of additional allelic

Table 1 Genetic instability syndromes

Disease*	OMIM†	Gene	Function	Major phenotypes
Werner syndrome	277,700	<i>WRN</i> (ref. 6)	Helicase/exonuclease	Skin atrophy, cataracts, diabetes mellitus, osteoporosis, hypogonadism, atherosclerosis, cancer predisposition ^{2,9}
Rothmund–Thomson syndrome	268,400	<i>RecQ4</i> (ref. 50)	Helicase	Poikiloderma, photosensitivity, skeletal abnormality, cataracts, cancer predisposition (osteosarcoma) ⁵¹
Cockayne syndrome, type A	216,400	<i>CKN1</i> (ref. 52)	WD repeat protein	Neurodegeneration, skeletal abnormality (widened face), impaired sexual development, photosensitivity ⁵⁴
Cockayne syndrome, type B	133,540	<i>ERCC6</i> (ref. 53)	Helicase	
Ataxia telangiectasia	208,900	<i>ATM</i> (ref. 55)	Kinase	Cerebellar dysfunction, sensitivity to ionizing radiation, cancer predisposition ⁵⁶
Nijmegen breakage syndrome (Ataxia-telangiectasia variant)	251,260	<i>NSB1</i> (ref. 57)	Unknown	Microcephaly, growth retardation, immunodeficiency, cancer predisposition; sensitivity to ionizing radiation ⁵⁸

*The listed disorders are all autosomal recessives.

†OMIM, Online Mendelian Inheritance in Man (ref. 49).

variants, including those with substantially greater repeat lengths, even among sibships. Such meiotic expansion is presumed to be the basis of the observation of ‘anticipation’, in which earlier ages of onset are observed in sequential generations. Thus, individuals with alleles having phenotypic effects that escape the force of natural selection may have progeny with alleles that impair reproductive fitness, leading to a decrease in carrier prevalence. But meiosis can also lead to repeat contractions. The situation is further complicated by the occurrence of stochastic somatic instability of repeat lengths. Moreover, there is now evidence that we are observing variable impacts upon at least two contiguous loci, as the cataracts may be due to haploinsufficiency of the downstream *Six5* locus³². Thus, the population dynamics of mutational load and the variability of phenotypic expression at that locus are complex.

Disorders of lipid and carbohydrate metabolism

Genetic aberrations in lipid metabolism can lead to premature coronary atherosclerosis, peripheral vascular disease and strokes, and numerous gene variants can increase vulnerability to type 2 diabetes mellitus. It is beyond the scope of this article to review this research or to give details of the numerous progeroid features of advanced diabetics. Glycation of proteins is thought to be one of the principal molecular mechanisms of ageing³³ — a further rationale for considering the various forms of diabetes as segmental progeroid syndromes. Here we will briefly review a particularly interesting recessive form of insulin-resistant diabetes, congenital generalized lipodystrophy (Seip–Berardinelli syndrome)³⁴. This disorder is associated with a profound generalized atrophy and agenesis or dystrophy of metabolically active adipose tissue, but with sparing of small depots such as the orbits, palms, soles and periarticular regions, where mechanical functions may be important³⁵. Other features include hyperinsulinaemia, hypertriglyceridaemia and fatty change of liver and spleen, retinopathy, cardiomyopathy, anginal pectoris, myocardial infarction, osteosclerosis, proteinuria and renal failure (probably from Kimmelstiel–Wilson disease)³⁴. There is also severe amyloidosis of the pancreatic islets³⁶. Patients may have variable degrees of cognitive impairment and dilated brain ventricles and basal cisternae³⁷. A remarkable feature is what has been referred to as an ‘anabolic syndrome’ — these young patients have voracious appetites and exhibit rapid pre-pubertal growth, advanced bone age, muscular hypertrophy and precocious development of genitalia³⁴. The numerous subsequent pathologies are such, however, that they rarely live beyond the third decade of life. We seem, therefore, to be dealing with a syndrome of partly accelerated development followed by an array of degenerative disorders, including diabetes and its sequelae. Reports of an entirely different aberration in carbohydrate metabolism, involving a β -glycosidase, may similarly impact primarily upon various domains of development^{38,39}.

A locus for congenital generalized lipodystrophy has been mapped to the long arm of chromosome 9 at band 34 (9q34), where there is a plausible candidate gene, *RXRA* (ref. 40). This gene codes

for the retinoid X receptor alpha, thought to be crucial in the regulation of adipocyte differentiation.

Hutchinson–Gilford Syndrome

A more common designation for this exceedingly rare disorder is ‘progeria of childhood’ or simply ‘progeria’⁴¹. This is unfortunate, as it communicates the view that this is simply accelerated ageing. It is certainly the case that these subjects have a remarkable phenotype, including atrophy and/or failure of development of subcutaneous fat (noted in neonates), alopecia (loss of hair), short stature, premature atherosclerosis and musculoskeletal abnormalities. The musculoskeletal abnormalities, however, are more consistent with developmental abnormalities, such as agenesis of all or part of the clavicles and joint dislocations and arthritis resulting from congenital dislocations. The short stature of normal ageing is acquired, but in HGS it is developmental. Many common features of ageing in the general population are not seen in HGS, such as cognitive decline, Alzheimer lesions, cataracts, presbycusis (age-related hearing loss) and presbyopia (age-related visual impairment). Because of the rarity of these cases (about one per million), the issue of susceptibility to neoplasia remains open. One of seven patients whose histories were available died of a chondrosarcoma of the chest wall at the age of 13 years⁴². A preliminary study of gene expression in cultured fibroblast-like cells from two HGS subjects gave evidence consistent with mitotic instability⁴³. This includes findings of shortened telomere lengths and diminished DNA repair in progeric fibroblasts^{44,45}. Thus, HGS may eventually join the group of genomic instability syndromes listed in Table 1. If so, it would further strengthen the linkage between aberrations in DNA transactions, atherosclerosis, cancer and ageing. The formal genetics of HGS remains to be established. But given the lack of evidence for parental consanguinity and the positive associations with advanced paternal age, it is likely to be an autosomal dominant syndrome.

Conclusions

There is the potential for considerable genetic variability in the expression of ageing in human subjects. No single gene, however, can accelerate all aspects of the complex senescent phenotype. This conclusion raises a paradox. On the one hand, we have some 60 years of research showing that a simple environmental manipulation, caloric restriction, can extend the life spans of mammals and slow the rates of progression of a large number of senescent phenotypes. This indicates a relatively simple mechanistic modulation of ageing, with a possible principal role for the modulation of oxidative damage to macromolecules^{46,47}. Evolutionary theory and phenotypic analysis of single-gene mutations, however, argue that a variety of different mechanisms modulate various processes of ageing. Further research will hopefully resolve this apparent paradox. That research should devote more attention to the determination of the genetic basis for unusual resistance to specific senescent phenotypes of human ageing⁴⁸. We might thereby uncover common mechanisms leading to slower rates of ageing in multiple physiological systems, thus

providing a rationale for interventions applicable to most individuals. Alternatively, we may find a large variety of pathways, including multiple mechanisms of senescence within a single physiological system. Such a result would require a large variety of interventions tailored to the needs of individual subjects. □

1. Rose, M. R. *Evolutionary Biology of Aging* (Oxford Univ. Press, Oxford, 1991).
2. Martin, G. M. *Genetic Effects on Aging* Vol. II (ed. Harrison, D. E.) 493–520 (Telford Press, Caldwell, NJ, 1990).
3. Grist, S. A., McCarron, M., Kutlaca, A., Turner, D. R. & Morley, A. A. In vivo human somatic mutation: frequency and spectrum with age. *Mutat. Res.* **266**, 189–196 (1992).
4. Martin, G. M. *et al.* Somatic mutations are frequent and increase with age in human kidney epithelial cells. *Hum. Mol. Genet.* **5**, 215–221 (1996).
5. Crow, J. F. Spontaneous mutation in man. *Mutat. Res.* **437**, 5–9 (1999).
6. Yu, C. E. *et al.* Positional cloning of the Werner's syndrome gene. *Science* **272**, 258–262 (1996).
7. Fukuchi, K., Martin, G. M. & Monnat, R. J. Jr Mutator phenotype of Werner syndrome is characterized by extensive deletions. *Proc. Natl Acad. Sci. USA* **86**, 5893–5897 (1989). [Published erratum appears in *Proc. Natl Acad. Sci. USA* **86**, 7994 (1989).]
8. Epstein, C. J., Martin, G. M., Schultz, A. L. & Motulsky, A. G. Werner's syndrome a review of its symptomatology, natural history, pathologic features, genetics and relationship to the natural aging process. *Med. Baltimore* **45**, 177–221 (1966).
9. Goto, M., Miller, R. W., Ishikawa, Y. & Sugano, H. Excess of rare cancers in Werner syndrome (adult progeria). *Cancer Epidemiol. Biomarkers Prev.* **5**, 239–246 (1996).
10. Gimbrone, M. A. Jr Endothelial dysfunction, hemodynamic forces, and atherosclerosis. *Thromb. Haemost.* **82**, 722–726 (1999).
11. Martin, G. M., Sprague, C. A. & Epstein, C. J. Replicative life-span of cultivated human cells. Effects of donor's age, tissue, and genotype. *Lab. Invest.* **23**, 86–92 (1970).
12. Ogburn, C. E. *et al.* An apoptosis-inducing genotoxin differentiates heterozygous carriers for Werner helicase mutations from wild-type and homozygous mutants. *Hum. Genet.* **101**, 121–125 (1997).
13. Castro, E. *et al.* Polymorphisms at the Werner locus: I. Newly identified polymorphisms, ethnic variability of 1367Cys/Arg, and its stability in a population of Finnish centenarians. *Am. J. Med. Genet.* **82**, 399–403 (1999).
14. Ye, L. *et al.* Association of a polymorphic variant of the Werner helicase gene with myocardial infarction in a Japanese population. *Am. J. Med. Genet.* **68**, 494–498 (1997). [Published erratum appears in *Am. J. Med. Genet.* **70**, 103 (1997).]
15. Huang, S. *et al.* The premature aging syndrome protein, WRN, is a 3' 5' exonuclease. *Nature Genet.* **20**, 114–116 (1998).
16. Shen, J. C. *et al.* Werner syndrome protein. I. DNA helicase and dna exonuclease reside on the same polypeptide. *J. Biol. Chem.* **273**, 34139–34144 (1998).
17. Suzuki, N., Shiratori, M., Goto, M. & Furuichi, Y. Werner syndrome helicase contains a 5' 3' exonuclease activity that digests DNA and RNA strands in DNA/DNA and RNA/DNA duplexes dependent on unwinding. *Nucleic Acids Res.* **27**, 2361–2368 (1999).
18. Poot, M., Hoehn, H., Runger, T. M. & Martin, G. M. Impaired S-phase transit of Werner syndrome cells expressed in lymphoblastoid cell lines. *Exp. Cell Res.* **202**, 267–273 (1992).
19. Yan, H., Chen, C. Y., Kobayashi, R. & Newport, J. Replication focus-forming activity 1 and the Werner syndrome gene product. *Nature Genet.* **19**, 375–378 (1998).
20. Lebel, M., Spillare, E. A., Harris, C. C. & Leder, P. The Werner syndrome gene product co-purifies with the DNA replication complex and interacts with PCNA and topoisomerase I. *J. Biol. Chem.* **274**, 37795–37799 (1999).
21. Gray, M. D. *et al.* The Werner syndrome protein is a DNA helicase. *Nature Genet.* **17**, 100–103 (1997).
22. Cooper, M. P. *et al.* Ku complex interacts with and stimulates the Werner protein. *Genes Dev.* **14**, 907–912 (2000).
23. Li, B. & Comai, L. Functional interaction between Ku and the Werner syndrome protein in DNA end processing. *J. Biol. Chem.* **275**, 28349–28352 (2000).
24. Balajee, A. S. *et al.* The Werner syndrome protein is involved in RNA polymerase II transcription. *Mol. Biol. Cell* **10**, 2655–2668 (1999).
25. Blander, G. *et al.* Physical and functional interaction between p53 and the Werner's syndrome protein. *J. Biol. Chem.* **274**, 29463–29469 (1999).
26. Hanada, K. *et al.* RecQ DNA helicase is a suppressor of illegitimate recombination in *Escherichia coli*. *Proc. Natl Acad. Sci. USA* **94**, 3860–3865 (1997).
27. Yamagata, K. *et al.* Bloom's and Werner's syndrome genes suppress hyperrecombination in yeast sgs1 mutant: implication for genomic instability in human diseases. *Proc. Natl Acad. Sci. USA* **95**, 8733–8738 (1998).
28. Fu, Y. H. *et al.* An unstable triplet repeat in a gene related to myotonic muscular dystrophy. *Science* **255**, 1256–1258 (1992).
29. Tiscornia, G. & Mahadevan, M. S. Myotonic dystrophy: the role of the CUG triplet repeats in splicing of a novel DMPK exon and altered cytoplasmic DMPK mRNA isoform ratios. *Mol. Cell* **5**, 959–967 (2000).
30. Harley, H. G. *et al.* Size of the unstable CTG repeat sequence in relation to phenotype and parental transmission in myotonic dystrophy. *Am. J. Hum. Genet.* **52**, 1164–1174 (1993).
31. Harnshere, M. G. *et al.* Myotonic dystrophy: the correlation of (CTG) repeat length in leucocytes with age at onset is significant only for patients with small expansions. *J. Med. Genet.* **36**, 59–61 (1999).
32. Klesert, T. R. *et al.* Mice deficient in Six5 develop cataracts: implications for myotonic dystrophy. *Nature Genet.* **25**, 105–109 (2000).
33. Lee, A. T. & Cerami, A. Role of glycation in aging. *Ann. NY Acad. Sci.* **663**, 63–70 (1992).
34. Seip, M. & Trygstad, O. Generalized lipodystrophy, congenital and acquired (lipoatrophy). *Acta Paediatr. Suppl.* **413**, 2–28 (1996).
35. Garg, A., Fleckenstein, J. L., Peshock, R. M. & Grundy, S. M. Peculiar distribution of adipose tissue in patients with congenital generalized lipodystrophy. *J. Clin. Endocrinol. Metab.* **75**, 358–361 (1992).
36. Garg, A., Chandalia, M. & Vuitch, F. Severe islet amyloidosis in congenital generalized lipodystrophy. *Diabetes Care* **19**, 28–31 (1996).
37. Westvik, J. Radiological features in generalized lipodystrophy. *Acta Paediatr. Suppl.* **413**, 44–51 (1996).
38. Kuro-o, M. *et al.* Mutation of the mouse klotho gene leads to a syndrome resembling ageing. *Nature* **390**, 45–51 (1997).
39. Martin, G. M. & Mian, I. S. New mice for old questions. *Nature* **390**, 18–19 (1997).
40. Garg, A. *et al.* A gene for congenital generalized lipodystrophy maps to human chromosome 9q34. *J. Clin. Endocrinol. Metab.* **84**, 3390–3394 (1999).
41. Brown, W. T. Progeria: a human-disease model of accelerated ageing. *Am. J. Clin. Nutr.* **55**, 1222S–1224S (1992).
42. Oshima, J., Brown, W. T. & Martin, G. M. No detectable mutations at Werner helicase locus in progeria. *Lancet* **348**, 1106 (1996).
43. Ly, D. H., Lockhart, D. J., Lerner, R. A. & Schultz, P. G. Mitotic misregulation and human ageing. *Science* **287**, 2486–2492 (2000).
44. Allsopp, R. C. *et al.* Telomere length predicts replicative capacity of human fibroblasts. *Proc. Natl Acad. Sci. USA* **89**, 10114–10118 (1992).
45. Wang, S., Nishigori, C., Yagi, T. & Takebe, H. Reduced DNA repair in progeria cells and effects of gamma-ray irradiation on UV-induced unscheduled DNA synthesis in normal and progeria cells. *Mutat. Res.* **256**, 59–66 (1991).
46. Sohal, R. S. & Weindruch, R. Oxidative stress, caloric restriction, and aging. *Science* **273**, 59–63 (1996).
47. Martin, G. M., Austad, S. N. & Johnson, T. E. Genetic analysis of aging: role of oxidative damage and environmental stresses. *Nature Genet.* **13**, 25–34 (1996).
48. Martin, G. M. in *Handbook of the Aging Brain* (eds Wang, E. & Snyder, D. S.), 126–134 (Academic, New York, 1998).
49. McKusick, V. A. *Mendelian Inheritance in Man*, 12th edn (Johns Hopkins Univ. Press, Baltimore, MD, 1998).
50. Kitao S. *et al.* Mutations in RECQL4 cause a subset of cases of Rothmund-Thomson syndrome. *Nature Genet.* **22**, 82–84 (1998).
51. Vennos, E. M. *et al.* Rothmund-Thomson syndrome: review of the world literature. *J. Am. Acad. Dermatol.* **7**, 750–762 (1992).
52. Henning, K. A. *et al.* The Cockayne syndrome group A gene encodes a WD repeat protein that interacts with CSB protein and a subunit of RNA polymerase II TFIIF. *Cell* **82**, 555–564 (1995).
53. Troelstra, C. *et al.* ERCC6, a member of a subfamily of putative helicases, is involved in Cockayne's syndrome and preferential repair of active genes. *Cell* **71**, 939–953 (1992).
54. Nance, M. A. & Berry, S. A. Cockayne syndrome: review of 140 cases. *Am. J. Med. Genet.* **42**, 68–84 (1992).
55. Savitsky, K. *et al.* A single ataxia telangiectasia gene with a product similar to PI-3 kinase. *Science* **268**, 1749–1753 (1995).
56. Gatti, R. A. *et al.* Ataxia-telangiectasia: an interdisciplinary approach to pathogenesis. *Medicine* **70**, 99–117 (1991).
57. Varon, R. *et al.* Nibrin, a novel DNA double-strand break repair protein, is mutated in Nijmegen breakage syndrome. *Cell* **93**, 467–476 (1998).
58. van der Burg, I. *et al.* Nijmegen breakage syndrome. *J. Med. Genet.* **33**, 20–32 (1996).

The future of ageing

Leonard Hayflick

Department of Anatomy, School of Medicine, University of California at San Francisco, PO Box 89, The Sea Ranch, California 95497, USA
(e-mail: len@gene.com)

Advances in our knowledge of age-associated diseases have far outpaced advances in our understanding of the fundamental ageing processes that underlie the vulnerability to these pathologies. If we are to increase human life expectancy beyond the fifteen-year limit that would result if today's leading causes of death were resolved, more attention must be paid to basic research on ageing. Determination of longevity must be distinguished from ageing to take us from the common question of why we age to a more revealing question that is rarely posed: why do we live as long as we do? But if the ability to intervene in ageing ever becomes a reality, it will be rife with unintended and undesirable consequences.

Research on ageing entered the main stream of biological inquiry about 30 years ago, but since then no notable advances have occurred in our understanding of the human ageing process. Success has been achieved only in our knowledge of age-related diseases.

The failure to distinguish between ageing research (biogerontology) and research on age-associated diseases (geriatric medicine) has been, and still is, a source of misunderstanding. And there is little evidence that this failure, with its important scientific, political and societal consequences, will soon be rectified. Thus, the present imbalance will continue, in which resources available for research on the diseases of old age far exceed those available to address the core question: why are old cells more vulnerable to disease than are young cells?

Policy-makers, properly impressed with the future demographics of the greying of all economically developed countries, are basing important policies and decisions on a flawed understanding of what constitutes ageing research and what they believe might be accomplished.

Disease and ageing

Ageing is not a disease and the distinction is central to an understanding of why the resolution of the leading causes of death in old age — cardiovascular disease, stroke and cancer — will tell us little about the fundamental biology of age changes. The resolution of all three conditions would result only in an increase of about 15 years in human life expectancy in the developed world¹, after which ageing will be revealed as the leading cause of death.

Disease processes can be distinguished from age changes for at least four criteria. Unlike any disease, age changes (1) occur in every animal that reaches a fixed size in adulthood; (2) take place in virtually all species; (3) occur in all members of a species only after the age of reproductive success; and (4) occur in animals removed from the wild and protected by humans even when that species has not experienced ageing for thousands or even millions of years.

The study of age-associated diseases and manipulation of development in lower life forms dominates what is mistakenly described as the field of ageing research. One example is that more than half the budget of the US National Institute on Ageing is spent on Alzheimer's disease research, yet motor vehicle accidents cause twice as many deaths¹ and from age 65 on, it is not even one of the five leading causes of death². The likelihood of dying from Alzheimer's disease is 0.7% (ref. 3) and the complete

Box 1

Ageing in feral animals

Ageing rarely if ever occurs in feral animals because it is unusual for them to live long enough to experience the phenomenon. The same observation can be made for prehistoric humans. Natural selection could not select for a process like ageing when few, if any, animals ever lived long enough to participate in the selection process.

If, through human intervention, feral animals are kept as pets or deposited in zoos and thus protected from predation, disease and accidents, age changes that may never have been experienced in the wild will be unmasked. The resulting greater longevity is not caused by the expression of new genes but by the protection provided by human intervention. The finding of old feral animals usually results from the enormous increase in the human population and the consequent disturbance of many ecological niches (for example, death of predators).

Most animals do not die immediately after reproductive success because it is prohibitively costly in terms of energy to evolve such a system. The class of animals generally referred to as 'big bang animals', represented by the Pacific salmon and the marsupial male rat, may seem to be an exception to this notion. However, it is more likely that the deaths that occur in these animals after reproductive success result from their unique expenditure of enormous amounts of energy that precedes mating⁵. Whether age changes occur is uncertain because the unusual rapidity of the events that precede death is unique and the fact that there is no necessity for death to be preceded by age changes.

resolution of this disease would add about 19 days onto average life expectancy¹.

In the minds of the public, policy-makers and many biomedical scientists, no one suffers or dies from ageing. We suffer and die from the diseases associated with the ageing process. But no one over the age of, say 75, has or will die from what is written on their death certificate. Death results from the inevitable increase in systemic molecular disorder that living long enough incurs. That disorder simply increases vulnerability to whatever was, or will be, diagnosed as the cause of death. There are multiple pathologies in older people and, because there are few autopsies and little research, the true cause of death is rarely known.

Box 2

How long will we live?

There is no evidence to support the many outrageous claims of extraordinary increase in human life expectancy that might occur in our lifetime or that of our children. Based on the US Census Bureau Middle Series, life expectancy in 2050 will be about 82 years for both sexes in the United States¹⁴. The US Social Security Administration anticipates a life expectancy of 78.1, 80.4 and 83.5 years for both sexes in 2066 based on three alternative assumptions about decreases in mortality rates¹⁵. The G7 industrialized countries project life expectancy at birth in 2050 to range from a high of 83.5 in France to a low of 80.5 in the United States¹⁶. In a more recent analysis, in which it was assumed that future decline in mortality rates will follow the exponential decline that has occurred during the past 50 years, Tuljapourkar *et al.* forecast that life expectancy at birth in 2050 will vary from a high of 90.9 in Japan to a low of 82.9 in the United States¹⁶.

More than 75% of all human deaths in developed countries now occur in those over the age of 75. If the causes of these deaths are resolved we will not become immortal but we will have revealed how death occurs in the absence of disease. What will be found is that the underlying cause of these deaths is the inexorable loss of physiological capacity in the cells of vital organs — the hallmark of ageing.

Ageing and longevity determination

If ageing research is to advance, it will not only be necessary to distinguish biogerontology from geriatric medicine but it will also be necessary to distinguish ageing from longevity determination.

Ageing is a stochastic process that occurs after reproductive maturation and results from the diminishing energy available to maintain molecular fidelity. This disorder has multiple aetiologies including damage by reactive oxygen species. Longevity determination, on the other hand, is not a random process. It is governed by the excess physiological capacity reached at the time of sexual maturation that, through natural selection, was achieved to better guarantee survival. For this reason, the question “Why do we live as long as we do?” might be more appropriate than “Why do we age?”.

Species survival depends on a sufficient number of members living long enough to reproduce and, if necessary, to raise progeny to independence. Natural selection favours animals that have greater survival skills and, especially, redundant physiological reserve in vital organs beyond the minimum needed to survive the damage that might be exacted by predators, disease, accidents or environmental extremes. The amount of excess physiological capacity, like the amount of redundancy engineered into space vehicles, provides the potential for continued function beyond the primary goal^{4,5}.

Energy is better spent on guarantying reproductive success than it is for increasing individual longevity. Consequently, age-weakened individuals living beyond reproductive success have diminishing value for the survival of a species and will be culled by natural selection. The genome that governs molecular synthesis and integrity from conception to sexual maturation is incapable of maintaining fidelity indefinitely. After reproductive success the energy available to maintain orderliness diminishes so that continued survival is determined only indirectly by the genome. Increasing systemic molecular disorder, or ageing, occurs in spite of the action of normal repair processes, because these too incur disorder.

Genes do not govern ageing

Ageing is not a programmed process governed directly by genes. Studies in lower animals that have led to the identification of genes that are involved in ageing have not shown a reversal or arrest of the

inexorable expression of molecular disorder that is the hallmark of ageing. Those studies are more accurately interpreted as showing that certain genes impact on longevity determination because the results alter physiological capacity and occur before the ageing process begins.

Further evidence that genes do not have a direct role in the ageing process is that individual animals do not age at the same rate nor are the patterns of age changes identical. This results in the variations found in age of death. The quantitative and qualitative disorder of age changes stand out in stark contrast to the ordered changes that occur during genetically driven embryogenesis and development. The variability in the manifestations of ageing differs greatly from animal to animal but the variability in developmental changes differs trivially. Humans from conception to adulthood are virtually identical in respect to the stages and timing of biological development but from about thirty on, age changes make humans much more heterogeneous. Just as a blueprint is vital for manufacturing a complex machine and contains no information to cause its ageing, the genome is vital for biological development but contains no instructions for ageing.

Longevity determination in higher animals has been a profoundly neglected area of research. One class of animals that may provide insight into the mechanisms determining longevity are those animals that do not reach a fixed size in adulthood, and age either undetectably slowly or not at all. Animals of this class include some tortoises, many sport and cold water deep-sea fish, some amphibians and the American lobster. Whether these animals age at all, and the reasons for this, remain to be determined. They are not immortal because, like animals that do age, there is a constant threat of disease, predation and accidents⁶ (see Box 1). The time is long overdue for more intense study of the phenomenon of negligible ageing. Telomerase expression, for example, which is a hallmark of immortal cells in tissue culture, has been found at extraordinarily high levels in the cells of negligibly ageing animals such as the American lobster and the rainbow trout^{7,8}.

Is tampering with the ageing process desirable?

Life expectancy is the average total number of years that a human expects to live. This is fundamentally different from life span, which is the maximum number of years that a human can live.

The human life span has remained unchanged for the past 100,000 years at about 125 years^{4,5}. What has changed is life expectancy at birth, which has increased in the United States and other developed countries from about 49 years in 1900 to about 76 years in 1997⁹ (and is projected to continue increasing; see Box 2). This 27-year increase in life expectancy is equivalent to the increase in life expectancy that occurred from the time of ancient Rome until the year 1900. The increase has been caused by the substantial elimination of infectious diseases that occur in youth through better hygiene and the discovery of antibiotics and vaccines (see Box 3). It is the chronic diseases — cardiovascular disease, stroke and cancer — that remain unresolved and that dominate today as the causes of death in the elderly. Twenty-one of the 27-year increase in life expectancy that occurred during the twentieth century took place during the first 70 years. Only a six-year increase in life expectancy has occurred in the past 27 years¹⁰.

To know what the future societal impact might be of a 15-year increase in life expectancy, one might consider the changes that have occurred from 1931 until the present, which spans a period of time in which an approximate 15-year increase in life expectancy has occurred¹¹. Of the many observations that could be made, three are: the increase in the proportion of older people, the greater time spent in frailty and dependency in old age, and the political and economic consequences that both have had¹².

Despite the likelihood that biological ageing is inexorable and inevitable, is the power to manipulate the human ageing process a desirable goal?

Box 3

Ageing as an artefact of civilization

Ageing is a phenomenon unique to the human species because it is a consequence of our advancing knowledge of hygiene and biomedicine. The resulting increase in the numbers of older people in developed countries is, to a large extent, an unintended consequence of these advances and an artefact of human civilization^{4,5,10}.

Humans, and the animals we choose to protect, are the only species in which large numbers experience ageing. Furthermore, old humans, or old animals, are not essential for the survival of any species, and there seems to be no selective advantage favouring the survival of old individuals. The evidence for this is that prehistoric human remains have never revealed individuals older than about 50 years of age, and humans had a life expectancy at birth of 30 years or less for more than 99.9% of the time that we have inhabited this planet.

Members of exotic feral animal species, who for millions of years have not experienced ageing (see Box 1), reveal those changes when protected by humans as pets or in zoos. It would be difficult to explain how evolution could have selected for a process like ageing that could be made to appear in all members of a species after, perhaps, millions of years of suppression.

Because modern humans, unlike feral animals, have learned how to escape death long after reproductive success, we have revealed a process that, teleologically, was never intended for us to experience.

The pill

As an exercise to explore what would occur if tampering with the ageing process became possible, one might imagine the simplest method: a pill that either stops or temporarily arrests human ageing. The first concern is that those involved in the discovery and the rich and powerful will have earliest, or depending on availability, even the only, access. It is questionable whether these would be the most important to be favoured first, or at all. Presumably, the pill would also become available to the antisocial killers, tyrants and those guilty of genocide along with those who contribute to, or benefit, human civilization.

Of the many predicaments that could be imagined, one is: when in life would one choose to take the pill? Before making the decision to administer it in youth one should be aware of the fact that many people in their seventies and later will say that this is the happiest time of their lives, and to have had their ageing processes arrested at an earlier age would have denied them the happiness of retirement, travel, freedom from child-rearing responsibilities and unlimited time to pursue their interests.

Furthermore, at what age would one choose to have one's ageing arrested or retarded if one had not yet arrived at a sufficiently mature age to reach an informed decision? Would we be able to reverse the ageing process to return to an age that, with hindsight, had been our happiest? Is it not likely that the circumstances that had contributed to an earlier period of perceived happiness would no longer exist should one be able to return to that age?

We interact with each other to a large extent as a function of our perceptions of relative age. The destruction of such interpersonal relationships would be likely to have enormous negative personal and societal consequences. One of the many bizarre scenarios that could be imagined would be to find families in which children who chose not to take an anti-ageing pill would find themselves to be biologically older than their parents, who did.

It is possible that replacing all old body parts with new ones might

circumvent ageing. But our brains could not be replaced without a loss of self-identity and memory. Scenarios in which the information content of an ageing brain could be uploaded onto a computer and then downloaded onto a new 'erased' brain remain the province of science fiction.

The goal of arresting the ageing process might be viewed in the same light that we view the arrest of our physical or mental development in childhood — as a serious pathology. We should also bear in mind that arresting ageing is not likely to be an attractive option for the substantial part of the world's population who find themselves poor, oppressed, sick or all three.

Would the least imperfect scenario be a future society in which everyone lived to their 100th birthday in good physical and mental health, then to die on the stroke of midnight?

Future possibilities

As yet, we know of no way in which the human ageing process can be slowed. Caloric restriction is a probable exception which, although observed in many species, has yet to be demonstrated conclusively in humans¹³. Even so, a near-starvation diet is unlikely to be acceptable to those of us who value quality of life above quantity of life.

It is likely that a natural increase in the human life span is presently occurring but so slowly that our ability to detect it will only be made after millennia of careful record keeping. This belief is based on persuasive evidence in the fossil record that indicates that the life spans of most animals increase as evolution proceeds⁵.

As some civilizations have, our society must learn that ageing and youth should be valued equally if for no other reason than the youth in developed countries have an excellent chance of experiencing the phenomenon that they may now hold in such low esteem. Then, the misplaced passion for cosmetic surgery, anti-ageing nostrums and similar snake oil remedies touted to arrest ageing will be recognized for what they truly are — at best, a cover-up for an irreversible and inexorable process and, at worst, a delusion and waste of money by the uninformed.

If the main goal of our biomedical research enterprises is to resolve causes of death, then every old person becomes a testimony to those successes. Biogerontologists have an obligation to emphasize that the goal of research on ageing is not to increase human longevity regardless of the consequences, but to increase active longevity free from disability and functional dependence. □

1. Anderson, R. N. *US Decennial Life Tables for 1989–91* Vol. 1, No. 4, 7–8 (National Center for Health Statistics, Hyattsville, MD, 1999).
2. Hobbs, F. B. & Damon, B. L. 65+ in the United States (US Bureau of the Census Current Population Reports, Special Studies P23-190) 3–9 (US Government Printing Office, Washington DC, 1996).
3. Anderson, R. N. *US Decennial Life Tables for 1989–91* Vol. 1, No. 4, 6 (National Center for Health Statistics, Hyattsville, MD, 1999).
4. Hayflick, L. How and why we age. *Exp. Gerontol.* **33**, 639–653 (1998).
5. Hayflick, L. *How and why we age* (Ballantine Books, New York, 1996).
6. Hayflick, L. The illusion of cell immortality. *Brit. J. Cancer* **83**, 841–846 (2000).
7. Klapper, W. *et al.* Longevity of lobsters is linked to ubiquitous telomerase expression. *FEBS Lett.* **439**, 143–146 (1998).
8. Klapper, W., Heidorn, K., Kuhne, K., Parwaresch, R. & Krupp, G. Telomerase activity in "immortal" fish. *FEBS Lett.* **434**, 409–412 (1998).
9. Anderson, R. N. *United States Life Tables* (1997 National Vital Statistics Reports Vol. 47, No. 28) (Centers for Disease Control and Prevention, National Center for Health Statistics, 1999).
10. Hayflick, L. New approaches to old age. *Nature* **403**, 365 (2000).
11. Anderson, R. N. *United States Life Tables* (1997 National Vital Statistics Reports Vol. 47, No. 28) 32–33 (Centers for Disease Control and Prevention, National Center for Health Statistics, 1999).
12. Crimmins, E. M., Hayward, M. D. & Saito, Y. Changing mortality and morbidity rates and the health status and life expectancy of the older population. *Demography* **31**, 159–175 (1994).
13. Roth, G. S., Ingram, D. K. & Lane, M. A. Calorie restriction in primates: will it work and how will we know? *J. Am. Geriatr. Soc.* **47**, 896–903 (1999).
14. United States Bureau of the Census. *Population Projections of the United States by Age, Sex, Race and Hispanic Origin: 1995 to 2050* (Current Population Reports P25-1130) (US Government Printing Office, Washington DC, 1996).
15. Wade, A. *Social Security Mortality Projections* (1991 Trustees Report) (United States Social Security Administration, Washington DC, 1991).
16. Tuljapourkar, S., Nan, L. & Boe, C. A universal pattern of mortality decline in the G7 countries. *Nature* **405**, 789–792 (2000).

Resistance movement

Immunology and immunochemicals as tools and for their own sake.

Cell biology and immunology

Bender MedSystems

www.bendermedsystems.com

Antibodies, bioreagents and assay kits

This new catalogue from Bender MedSystems is organized in four sections listing products for use in cell adhesion, apoptosis, cytokines and tumour biology work. Antibodies, bioreagents and ready-to-use assay kits are listed, along with technical references and product specifications. A reference section offers technical information on methods for working with antibodies, immunochemical staining, flow cytometry, ELISA and whole blood assays.

Reader Service No. 100

ELISA conjugates and substrates

KPL

www.kpl.com

Ready-to-use formulations

Kirkegaard & Perry Laboratories produce antibodies labelled with peroxidase, phosphatase, biotin and fluorochrome to a wide range of animal species. The company also supplies ABTS and TMB HRP substrates in a ready-to-use liquid form. BluePhos AP substrate is stabilized as a two-component liquid that offers greater sensitivity than pNPP. Requests for standing orders and bulk packaging are welcomed.

Reader Service No. 101

Sci-70 antigen

Biodesign International www.biodesign.com

Autoimmune reagents

The presence of autoantibodies against the Sci-70 antigen is used as a specific serologic marker in research into systemic sclerosis. Biodesign's purified Sci-70 antigen's homogeneity is verified by SDS page. Multiple preparations of Jo-1, RNP-Sm, Sm, SS-B, SS-A, pANCA, cANCA, and TPO antigens for building autoimmune assays are also available, as is a companion plate-coating stabilizer in liquid or lyophilized form. Applications include standard solid-phase immunoassays.

Reader Service No. 102

QuantiBRITE

Becton Dickinson

www.bdfacs.com

Quantitative flow cytometric assay

New from BD Biosciences, the QuantiBRITE Anti-HLA-DR PE/Anti-Monocyte PerCP-Cy5.5 reagent was developed for measuring monocytic human leukocyte antigen expression in flow cytometry, in under an hour. It has been reported that the level of monocyte



Sectional interest: Bender's new catalogue.

class II expression might identify patients who are at high risk of developing infectious complications following trauma.

Reader Service No. 103

Cancer antigen antibodies

Zymed

www.zymed.com

Tumour-specific markers

Eight new monoclonal antibodies have been developed to help detect four high molecular weight glycoproteins — CA 15-3, CA 19-9, CA 125 and CA 72.4 — in order to aid in the diagnosis and monitoring of malignant tumours. These antibodies are relatively tumour-specific: CA 15-3 for breast cancer, CA 19-9 for colorectal and pancreatic cancer, CA 125 for non-mucinous ovarian cancer and CA 72.4 for a variety of human adenocarcinomas.

Reader Service No. 104

Mouse cytokine fusion proteins

Alexis Biochemicals

www.lexis-corp.com

Longer circulating half-life

These fusion proteins combine the Fc region from mouse antibody with a cytokine, or the ectodomain of a cell surface receptor or adhesion molecule, resulting in hybrid molecules with the properties of the antibody Fc domain as well as the ligand of interest. The mouse cytokine fusion proteins include non-lytic IL-2:Fc, IL-4:Fc and IL-10:Fc. Unlike most cytokines, which have short circulating half-lives, IL-10:Fc has a half-life of 31 hours. The proteins achieve constant circulating levels without continuous administration *in vivo*.

Reader Service No. 105

Antibodies at a discount

StressGen

www.stressgen.com/fallspecial

Hurry, hurry

Until 31 December 2000 (or while stocks last) StressGen is offering over 100 products at half-price or less, including antibodies and proteins in the areas of signal transduction, apoptosis, neurobiology, and associated accessories. Visit the company's website to see what's on offer.

Reader Service No. 106

BoneTRAP

Axis-Shield

www.axis-shield.com

High-specificity bone resorption measurement

The BoneTRAP immunoassay is claimed to improve monitoring of bone disease by providing rapid, specific determination of bone resorption rate. Using a solid-phase immunofixed enzyme activity assay, BoneTRAP measures serum levels of active, osteoclast-derived TRAP 5b, a specific marker secreted into the circulation during bone resorption. The assay detects only active TRAP molecules freshly liberated into the circulation without

ADVERTISEMENTS

Diagnostic Reagents

Pefachrome®-Series: chromogenic and fluorogenic substrates

Pefabloc®-Series: serine proteinase inhibitors

Purified snake venom proteins (e.g. Protac®, Ecarin, Batroxobin)

PENTAPHARM LTD

CH 4002 Basel Fax: +41 61 319 96 19
<http://www.pentapharm.ch>

READER ENQUIRY NO. 16

READER ENQUIRY NO. 33

new on the market

interference from inactive TRAP molecules or fragments. Suggested uses include monitoring antiresorptive treatment of bone disorders such as osteoporosis and Paget's disease. Reader Service No. 107

EZ Answers brochure

Pierce www.piercenet.com
Antibody FAQs

Pierce is offering researchers working with antibodies a free 40-page brochure that answers the most frequently asked questions related to antibody production, antibody purification, antibody modification and assay development and includes product comparison tables. Readers are offered a free EZ-Link biotinylation reagent selection guide poster. Reader Service No. 108

ORF-73 ELISA kit

ABI www.abionline.com
New test for human herpesvirus-8

Advanced Biotechnologies has introduced a new ELISA kit to detect latent IgG antibody to HHV-8. This virus is implicated in Kaposi's sarcoma, multicentric Castleman's disease, primary effusion lymphoma, reactive lymphadenopathy Bowen's disease and pemphigus vulgaris with or without HIV infection. Results obtained with the ORF-73 kit are said to compare favourably with those achieved using the existing indirect immunofluorescence assay (IFA-LNA), with a concordance of >96% between the two tests, and evidence of greater sensitivity using ORF-73. Reader Service No. 109

Sepharose 4 Fast Flow

Amersham Pharmacia Biotech
www.apbiotech.com

Immunoprecipitation for beginners

This is a starter pack for immunoprecipitation containing protein A and protein G Sepharose 4 Fast Flow with step-by-step instructions, buffer recipes, a troubleshooting guide and



On your marks: immunoprecipitation for all.

references. The two complementary binding selectivities isolate antibodies from various crude cell extracts, bind to a wide range of species and remove immune complexes formed in the immunoprecipitation reaction. They offer low ligand leakage and minimal non-specific adsorption to give minimum background with SDS-PAGE. Reader Service No. 110

Matchmaker

Clontech www.clontech.co.uk
Co-immunoprecipitation kit

This Co-IP kit contains haemagglutinin and c-Myc antibodies used to immunoprecipitate epitope tagged proteins. Interacting proteins are co-immunoprecipitated with the epitope tagged protein. Detecting protein interactions with each of these antibodies further confirms that the proteins under investigation do indeed interact. For researchers using older Matchmaker Gal4-based vectors, the Co-IP primer sets included in the kit can be used to amplify inserts so that they contain T7 promoters and epitope tags. Reader Service No. 111

Custom antibody service

Quantum Biotechnologies www.qbi.com
MAP technology

Quantum offers a full line of custom antibody services including production of polyclonal/monoclonal and bifunctional antibodies,

purification, conjugation, sequence analysis and monoclonal antibody modelling. Quantum claims that its *in vitro* technology produces monoclonal antibodies that are purer and cost less than ascites. The company can supply quantities from 10 mg to 100 g and the methodology is suitable for all hybridomas. Multiple antigen peptide to antibody technology is used to construct a highly immunogenic molecule that eliminates the need for carrier protein. Reader Service No. 112

Hybridoma detection

Marin Biologic Laboratories
www.marinbio.com

Identify rare antibodies

This detection technique uses encapsulated hybrid cells from immunized animals. Each hybrid cell is analysed for monoclonal antibody production by trapping the antibody in the encapsulated matrix. Cells with positive production are cloned and expanded. Purified monoclonal antibody can then be manufactured from milligram to multigram levels. Reader Service No. 113

Anti-c-myc-Eu(K) and Anti-c-myc-XL 665

Packard Instrument Company
www.packardinst.com

Reagents for HTRF assays

Two new epitope tag recognition reagents have been released for assay development and high-throughput screening. Anti-c-myc-Eu(K) is mouse monoclonal antibody conjugated with europium cryptate, the primary fluorophore of homogeneous time-resolved fluorescence chemistry. Anti-c-myc-XL 665 is the same antibody conjugated with the secondary fluorophore of HTRF, XL665, crosslinked allophycocyanin. Suggested applications are high-throughput screening of protein/protein interactions and receptor/ligand binding assays. Reader Service No. 114

These notes are compiled in the Nature office from information provided by the manufacturers.

ADVERTISEMENTS

Silver Stain Kit II for Electrophoresis

Rapid, sensitive detection
of proteins & nucleic acids

Wako

Wako BioProducts
Wako Chemicals USA, Inc.
1600 Bellwood Road
Richmond, VA 23237
Telephone: (804) 271-7677
Facsimile: (804) 271-7791
(800) 992-9256
bioproducts@wakousa.com

BACs • YACs • PACs cDNAs

Research Genetics provides clones and colony membranes for a number of human, mouse and rat libraries as well as a custom screening service for the libraries.

Research Genetics

U.S. or Canada 800-533-4363
U.K. 0-800-89-1393
FAX 256-536-9016

Check out our homepage at
<http://www.resgen.com>

Database Management Systems



Bioinformatics
Powered by GCG

- Comprehensive analysis packages
 - Enterprise-wide solutions
 - Extendable software
 - Computing security
 - Scriptable analysis

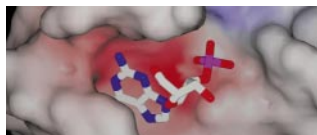
www.gcg.com

GCG is an Oxford Molecular Company

Client/Server & Desktop Software

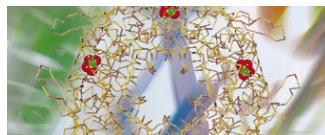
READER ENQUIRY NO. 41

READER ENQUIRY NO. 36



Determined effort

Structural genomics will enable scientists to make good use of DNA sequence data



On automatic

By setting up protein structure factories, structural biologists hope to speed up drug design

Structural genomics — from cottage industry to industrial revolution

As the new discipline of structural genomics takes off, the demand for a wide variety of skilled workers is set to rise, says Diane Gershon.

A common complaint among pharmaceutical companies is that they are target-rich and structure-poor. New initiatives — both public and private — in structural genomics, which aim to solve protein structures at previously unprecedented rates, should change all that (see page 130 of this issue and the supplement to this month's *Nature Structural Biology*).

The emerging field of structural genomics is at the interface of many disciplines. In addition to experts in protein expression, protein engineering, protein purification, structure determination and analysis, it also calls for engineers to develop robotic, high-throughput systems, and information-technology specialists to manage lab systems. The traditional academic approach to solving structures — having a graduate student go all the way from raw DNA data to structure determination and interpretation — is no longer appropriate. Structural genomics requires new ways of working that many — particularly those in the academic community — are not used to.

Yet the nature of the work does raise career issues. Many of the stages along the production line will eventually be done by robots with — it is hoped — minimal human intervention. Consequently, there is some concern that postdocs and graduate students could end up doing the work of technicians and lab assistants. Universities and funding agencies are sensitive to this and “don’t want students just turning a crank”, says Gaetano

Montelione of Rutgers University. “I think there’s enough technology and methodology development and fundamental science that it does make sense as a university activity.”

Gaetano is head of the Northeast Structural Genomics Consortium, one of seven new pilot centres in the United States set up with fund-

ing from the National Institute of General Medical Sciences (NIGMS). In fact, the NIGMS strongly discourages the use of its programme as a training vehicle for postdocs and grad students. This should be “very limited and needs to be justified”, says John Norvell, programme director in the institute’s division of cell biology and biophysics.

With the NIGMS initiative, the emphasis is squarely on technology development. Investigators must seek other sources of funding to springboard into other areas, such as the functional characterization of proteins. “I think a lot of us wanted to put more emphasis on the biology,” says Ian Wilson of the Scripps Research Institute in La Jolla, California. Wilson is principal investigator of another of the NIGMS’s pilot centres, the Joint Center for Structural Genomics.

Other initiatives are taking a more integrated approach from the start. Rather than targeting structural genomics specifically, the Wellcome Trust in Britain has, for example, just launched a £30 million (US\$43.5 million) functional genomics initiative that leaves it up to the applicants to identify their own scientific objectives.

Job prospects

There are jobs to be had in structural genomics, with start-up companies recruiting heavily. The demand for bioinformaticians still outstrips supply, although the shortage is not limited to structural genomics. “There’s a shortage of people who have very good training in production-level protein expression and purification,” says Montelione. “Somehow this was not viewed as important or valuable science, but getting samples that behave well, that crystallize and give good spectra is a big challenge,” he says.

Another area where Montelione sees potential shortages is enzymology. But traditional enzymology has been played down relative to molecular biology over the past 10–15 years, he says. “So the pipeline for enzymologists ... is not as good as it needs to be.” The real goal here is to understand the



Wilson: wanted more emphasis on biology.

function of the gene products. “At the end of the day, you’re going to have to do some kind of enzymology to validate the biochemical function of proteins,” Montelione says.

Playing the numbers game

A handful of start-up companies have formed in the past year or so, some of which have sparked considerable interest among venture investors. “The explosion in business concepts and new ideas [in post-genomics] really started in the fourth quarter of 1999,” says Samuel Colella, managing director of Versant Ventures in Menlo Park, California. The company was formed a year ago to seed healthcare-focused companies from an investment fund of \$250 million. Colella says he receives 30 e-mails a week from budding entrepreneurs and expects Versant to see close to 2,000 business plans this year alone.

The biotechnology industry experienced a bit of a lull in the late 1990s, says Colella, but publicity surrounding the completion of the draft sequence of the human genome has energized both the scientific and investment communities. Most importantly for Colella, it has been the quality of the business plans and the scientific and management credentials of key personnel in some of these start-ups that have caused him to look twice. “The class of 1999/2000 is a better class,” he says. Ten years ago, most entrepreneurs came out of academia or pharmaceutical companies; today, many more are from biotechnology



Montelione: notes possible labour shortage.

companies and “so are seasoned”, he says. “They’ve commercialized technology and lived through the wars.” Because of this, Colella says Versant Ventures has done more deals this year than it had planned.

In the structural genomics area, Versant Ventures has invested in start-up Syrrx in La Jolla, California. Syrrx is a spin-off from the non-profit Genomics Institute of the Novartis Research Foundation and has so far raised about \$25 million in two rounds of venture financing. “We capture any and all of the technologies from the [Novartis] institute in the areas of cloning, protein production, protein crystallization, collection of data. We have exclusive worldwide rights to this to use it within the field of structural biology,” says Nathaniel David, co-founder and director of business development at Syrrx. The company now employs just over 20 people, but David expects this to rise to 100 or so over the next couple of years.

“Structure-based drug design in pharma has been there, it’s been accepted over the past decade, but it’s not been highly invested in,” says Wendell Wierenga, Syrrx’s chief executive officer. Wierenga is a veteran of the pharmaceutical industry with more than 25 years of experience. But pharmaceutical companies are now beginning to realize that structural genomics “is an enabling technology if it can be managed and if it can be scaled to the levels that we’re talking about”, he says.

With what the company calls one Syrrx unit of around 65 people, David says it hopes to produce about 180 new structures per year. “But we’re going to be able to scale that process as a function of what really is valuable to the company,” he says. “We’re not in the business of selling databases. We don’t intend to be.” That said, companies such as Incyte Genomics of Palo Alto, California, have built successful businesses around just such a model.

“It may be that what makes sense for us is not to scale our ability to solve structures at an ever-growing rate but rather understand how to use this high-throughput structural

information to produce drugs at a historically unprecedented rate,” says David. Although Syrrx plans to forge partnerships with pharmaceutical companies in the near term, Wierenga expects the business to evolve and develop its own drug discovery effort. As a step towards this, the company is about to announce a deal with MolSoft of La Jolla, which will provide Syrrx with the exclusive rights to its computational docking algorithm for use in structural genomics.

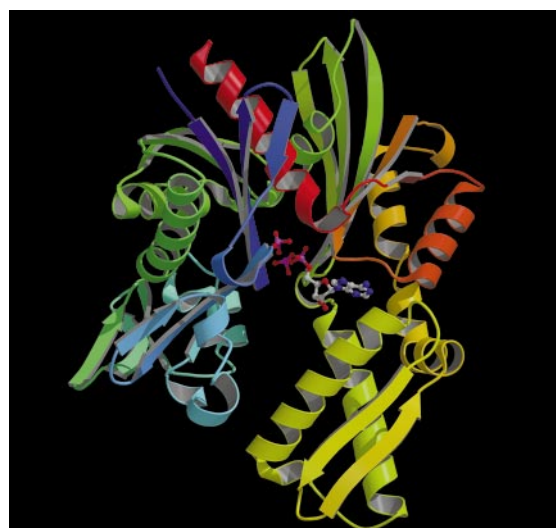
Another company riding the recent wave of investor interest is Structural GenomiX, located down the road from Syrrx. In just 14 months, the company has raised \$85 million in three rounds of venture-capital financing. Structural GenomiX seems more likely to pursue the Incyte business model by amassing vast quantities of structural data and selling subscriptions to its private databases to industry. As such, the company hopes that its healthy bank balance will allow it to get a head-start on the competition. “We’re expressing thousands of proteins, but we’re not determining the structure of thousands of proteins today,” says Tim Harris, president and chief executive officer of Structural GenomiX. “We will be doing hundreds in due course.”

Harris is a co-founder of the company and has 20 years of experience in the industry under his belt. Before forming Structural GenomiX he was senior vice-president for R&D at Sequana Therapeutics (now Axys Pharmaceuticals); before that, he was director of biotechnology at Glaxo. Like Syrrx, the company has grown considerably in a short period. Since March last year it has gone from two to 65 employees, and Harris expects this to double by this time next year. Because of automation, he anticipates that more people will be hired in the docking and molecular modelling area — the “adding value area”, as Harris calls it.

A hard road to hoe

Nevertheless, the path to profitability is by no means straightforward. Producing protein structures on an industrial scale is a capital-intensive endeavour and presents a more formidable challenge than high-throughput DNA sequencing. “There are technical challenges at every step of the process,” says Harris, which is why he says he adopted the strategy of raising as much money as possible up front.

Both companies use X-ray crystallography to solve structures, which involves using synchrotron radiation sources to generate the powerful X-rays.



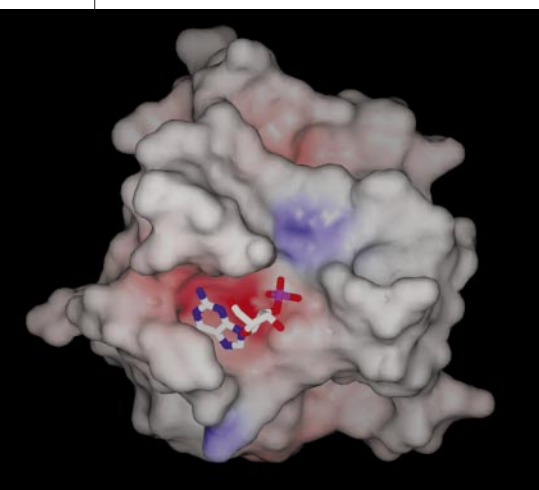
Structural genomics promises to speed up the discovery of protein structures.

Although beam time at synchrotrons is not limiting at the moment, says Harris, time on lines dedicated to protein structure determination is “definitely going to be at a premium”. In July, Structural GenomiX signed an agreement with the Advanced Photon Source at the US Department of Energy’s Argonne National Laboratory near Chicago, under which it will build a dual-undulator beam-line facility there solely for the company’s use. The facility will be ready in 14–18 months. Syrrx has also bought a beam line, which became functional last month at Lawrence Berkeley National Laboratory’s Advanced Light Source facility.

Despite the technical hurdles, both companies have strong credentials. Structural GenomiX’s scientific founders are noted crystallographer Wayne Hendrickson and Barry Honig, who uses theoretical and computational tools to study macromolecular structure and function. In addition to David, Syrrx’s co-founders include Peter Schultz, director of the Novartis institute, and Raymond Stevens, a professor in the molecular biology department at Scripps.

Neither Colella nor the companies see much conflict between the ongoing public and private efforts in structural genomics. Both companies plan to protect proprietary technology and targets, although, as it is also a member of the Joint Center for Structural Genomics, Syrrx will have to make public any contribution it makes on behalf of the centre. “I think the consortiums are going to try to do the entire protein landscape whereas [at Syrrx] we’re going to focus on areas of medical interest,” says Colella, who is chairman of the company’s board of directors. “There will be an element of overlap but we don’t consider it to be too big or too much of a concern,” says Harris. “I think there’s plenty to go round.”

Diane Gershon is Assistant Editor, New Technology, Nature Medicine.



In action: a compound bound to a protein.